paper 1

by Yasser Djawad

Submission date: 24-Feb-2021 12:12PM (UTC+0500)
Submission ID: 1516869536
File name: IIS25.03_03.pdf (1.43M)
Word count: 3761
Character count: 19962

IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Ingénierie des Systèmes d'Information
Vol. 25, No. 3, June, 2020, pp. 303-309
Journal homepage: http://iieta.org/journals/isi

Discrimination of Nitrogen Concentration of Fertilized Corn with Extracted Algae and Polymer Based on Its Leaf Color Images

Yasser Abd Djawad[1*], Hamza Rehman[2], Oslan Jumadi[3], Muhammad Tufail[2], Shahzad Anwar[2], Nathalie Bourgougnon[4]

[1] Department of Electronics, Engineering Faculty, Universitas Negeri Makassar, Makassar 90224, Indonesia
[2] Department of Mechatronics Engineering, University of Engineering and Technology, Peshawar 25000, Pakistan
[3] Department of Biology, Mathematics and Sciences Faculty, Universitas Negeri Makassar, Makassar 90224, Indonesia
[4] Laboratorie de Biotechnologie et Chimie Marines (LBCM), Universite de Bretagne-Sud, Vannes 56017, France

Corresponding Author Email: yasser.djawad@unm.ac.id

https://doi.org/10.18280/isi.250303

Received: 3 April 2020
Accepted: 5 June 2020

Keywords:
nitrogen level, colour intensity, image processing

ABSTRACT

Determination of nitrogen levels in plants is essential for variable rate fertilizer application in precision agriculture. In the past, several techniques have been developed for nitrogen concentration estimation in plants and crops employing vision system, however, they are computationally expensive and hence requires a considerable amount of time to produce accurate results. The technique developed in this work the determination of nitrogen levels in plants could be achieved effectively in real-time time by advance image processing techniques, machine visions and support vector machine (SVM) with MATLAB. The developed technique processes leaf's colored image via examining it Red, Green and Blue (RGB) values and compares them with standard intensity levels. The experimental results show effectiveness of the developed technique and accurately detect low or high concentration levels in corn. In addition, this method depends on two techniques for a final result, i.e. color intensity and SVM. If the answer is not similar between the two techniques the process will be repeated until the detection is similar. This study could be applied to a variety of crops, since this technique does not require large collection of data for training and special expertise for its on-field application.

1. INTRODUCTION

Crops growth is influenced by two variables, namely genetic variables and environmental variables. Nutrients availability in soil is one of the key environmental variable and Nitrogen (N) is one of important nutrient for growth. N-content in crops plays an important role in photosynthesis and can increase crops yield. However, excess N elements in crops could cause a number of problems too, such as plant collapse due to large leaf size, increased production costs and environmental problems due to N_2O gas emissions [1, 2].

Corn (Zea Mays L.), like other crops, also requires nitrogen nutrients for its growth. To provide additional nitrogen nutrients, nitrogen fertilizer is applied. Nitrogen provision as fertilizer is based on soil condition and crops need. In order to measure the nutrient content of crops, Soil Plant Analysis Development (SPAD) is employed by looking at the chlorophyll content of the leaves. Furthermore, researchers incorporate an alternative method by observing the leaf colour. Leaf colour could possess information related to level of N fertilization for a crop [3-6].

Various crops have been targeted in literature incorporating vision based techniques. Deficiency and excess of N in soybean leaf were investigated using the RGB and HSI models [3]. Similarly, experiments have conducted using various methods such Fourier and Wavelet transforms to extract colour and texture of the tomato leaves [7]. Corn leaf has been investigated using Dark Green Colour Index (DGCI) [8]. Nitrogen status of rice leaf was studied using segmentation technique [5].

Plants defects were examined using pattern recognition techniques utilizing leaf surface texture [9, 10]. In this study we move a step ahead from determining defects and plants disease to determining the level of fertility of a plant by examining its nitrogen level [11]. The proposed algorithm overcomes the traditional way of manually examining plants and uses an automated system using image processing techniques. The following four techniques are widely used for examination of nitrogen in plants [12].

Chemical Test method: determines nitrogen in plants by the use of chemical analysis which is mostly time consuming and expensive [9]. The method requires expensive chemicals and a team of trained laboratory technician to perform experimentation, after which an expert would rule out the result. This process takes up weeks and therefore the plant condition may have changed while the test is being performed.

SPAD Meter: measures transmission of light through leaves and has been proved as an industry standard for measurement [10, 11, 13, 14]. However, there are several factors that may compromise the measurement accuracy such as plant growth stage, environment condition, and presence of pests etc. [15]. Despite its good result the method has a limitation, i.e. very time consuming expensive and depends on many external factors.

Normalized Difference Vegetation Index (NVDI): method could be used to measure plant activity such as photosynthesis which also leads to chlorophyll and nitrogen content. This technique is based on measurements taken through a satellite

system [16, 17]. This makes the system expensive, complicated and complex.

Leaf color chart: has replaced the SPAD chlorophyll meter for plant nitrogen level examination. This technique is simple and inexpensive to use and uses a strip that contains color from yellowish to dark green. There are generally four or more color shades [18]. The green chart is usually used as a reference state of the plant showing low or high concentration as shown in Figure 1. This method involves the use of computer vision and automated techniques. This method saves the cost and effort involved in sending experts to manually observe the plant, take samples and send it off to the laborites for further testing. to come to the concussion of the hydrogen level which is very time consuming and expensive, as many farmers may not be able to afford a botany expert and lab tests. During the process, chances are, that the plans may have bad effect due to imbalance of nitrogen thus an automated technique will give an edge to the previous old techniques in identifying the level of nitrogen. Therefore, this method was used in this study.

Figure 1. Leaf color chart

2. MATERIALS AND METHODS

2.1 Data collection and design of experiment

The field experiment site was laid in the Cereal Crops Research Institute, Maros, South Sulawesi, Indonesia (4°58'55.3"S, 119°34'28.2"E). The area is located in a tropical climate zone. Climate is classified as Af based on the Köppen-Geiger climate classification system. Average temperature is 29°C and average rainfall is 347 mm/month. The soil at the experiment site was an alluvial soil type. The texture of field soil was 8% sand, 54% silt, 38% clay which belongs to texture classified as silty clay loam [19].

The site for experiments was divided into two corn fields. One field was fertilized with extracted algae (Algae field) and another field was fertilized with polymer (Polymer field). Algae field consisted of the following characteristic: (1) controlled corn that did not receive any treatment with any fertilizer, (2) corn fertilized using NPK, and (3) corn fertilized using 30% extracted algae; 60% extracted algae; and 90% extracted algae. Polymer field consisted of: (1) controlled corn that did not receive any treatment with any fertilizer, (2) corn fertilized using NPK, and (3) corn fertilized using polymer (without zeolite), polymer zeolite 15% and polymer zeolite 30%. The amount of N was 150 kg/ hectare and P and K were 50 kg/hectare respectively. Fertilizer was given once at the beginning of planting. However, in this study focused on leaves that are fertilized with algae and polymer. Three months after plantation, leaves were cut and gathered for each sample. Leaves were cut into three sections: base of leaf, middle leaf and leaf shoots. The leaves were scanned using a Canon L120 having resolution of 400 dpi and were saved in PNG (Portable Network Graphics) file-format.

The most important experiments conducted was on the unsupervised machine learning algorithm, Support Vector Machine, for training and teaching the algorithm to accuracy discriminate between high and low level of nitrogen. For this purposed, the scanned images were stored to create a database. The database provided with verity data points and training points for the classifier to train on. The three layers, i.e. based on leaf, middle of leaf and leaf shoots were used to determine the accuracy of the machine learning technique. This section of experiments was conducting in three section, first to determining level of nitrogen in the base of the leaf, then the middle and in the end the leaf shot. Since these images were provided in section, it was compared if all three section were classified in the same group or different group. The classier was tested and experimented on nine thousand images in the experimented site before it could be used outside the experiment filed.

Subsequent to a detailed review about slandered accuracy rate, the classifier was used to determine new samples. This step was performed to evaluate the accuracy of the SVM for new and unfamiliar cases. The SVM proved to properly identify new cases and unfamiliar cases. The new and unfamiliar cases were stored in the database, in this way the database will have new data points to train on and the evaluation process will be updated and more robust. The final result is compared with the universal chart and the answer feature by the classifier.

2.2 Description of the image processing algorithm

Previously developed methods have been proven to be time consuming and expensive, an automated method that uses non-intrusive computer vision techniques has always been a better alternative.

For performing computer vision and image processing techniques, MATLAB R2019 was used. The most basic reason for choosing this software was based on its strong and powerful control and image processing toolbox and solving any problem by convert data into matrix form and then performing operation on them.

The PNG images recorded previously provided raw data and was processed employing a tool. Features and points were extracted from the input images. The pre-processing of color scanned images included: (i) smoothing of images, (ii) noise removal, and (iii) image pixilation. The reason to pixelate images is to reduce extra load on the processor ensuring its real-time performance and faster response. Mathematical equations (as listed following), were introduced for calculation for the process of feature extraction for improved processing. These features are enormous importance for the classifier to properly detect the level of nitrogen in an image. Eq. (1) was used for smoothing the image.

$$c(x,y) = \frac{1}{K}\Sigma_{(s,t)\in S_{x,y}} c(s,t) \quad (1)$$

Here K is number of pixel in neighborhood having a center of x,y and $S_{x,y}$ defines set of coordinates defining neighborhood $K(x,y)$. As the images provide were RGB, three versions of the equations were generated,

$$c(x,y) = \frac{1}{K}\Sigma_{(s,t)\in S_{x,y}} R(s,t) \quad (2)$$

$$c(x,y) = \frac{1}{K}\Sigma_{(s,t)\in S_{x,y}} G(s,t) \quad (3)$$

$$c(x,y) = \frac{1}{K}\Sigma_{(s,t)\in S_{x,y}} B(s,t) \quad (4)$$

where, R, G, B stand for Red, Green, Blue, respectively. Subsequent to color intensity detection, a machine learning classifier is also implemented for determining the state of the plant. For this, a dataset was generated and then provided to unsupervised learning classifier, support vector machine, for detecting the state of the plant. The SVM worked on the dataset to train on data points. The SVM works fast and accurately due to the kernel trick, which are the option known for pattern recognition. If the universal chart and the classifier do not give the same answer, then the process will repeat until both answers are similar. This process ensures accurate answer.

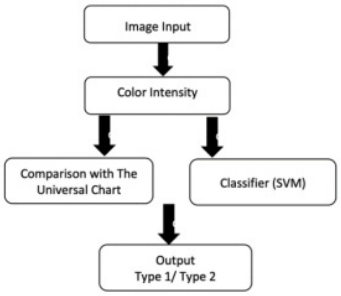
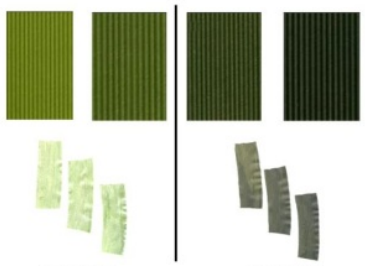


Figure 2. Experimentation flowchart

Figure 2 shows the chart used for nitrogen detection to classify plants with high (Type 2) or low (Type 1) nitrogen level. Results of classification are given in Figure 3. The experimental results show that the color intensity levels (each having a unique however similar pattern) have enough information to set a base for further development. Variation in color intensity can be related to the changes in nitrogen levels. Light green shows lack of nitrogen whereas dark green shows access of nitrogen.



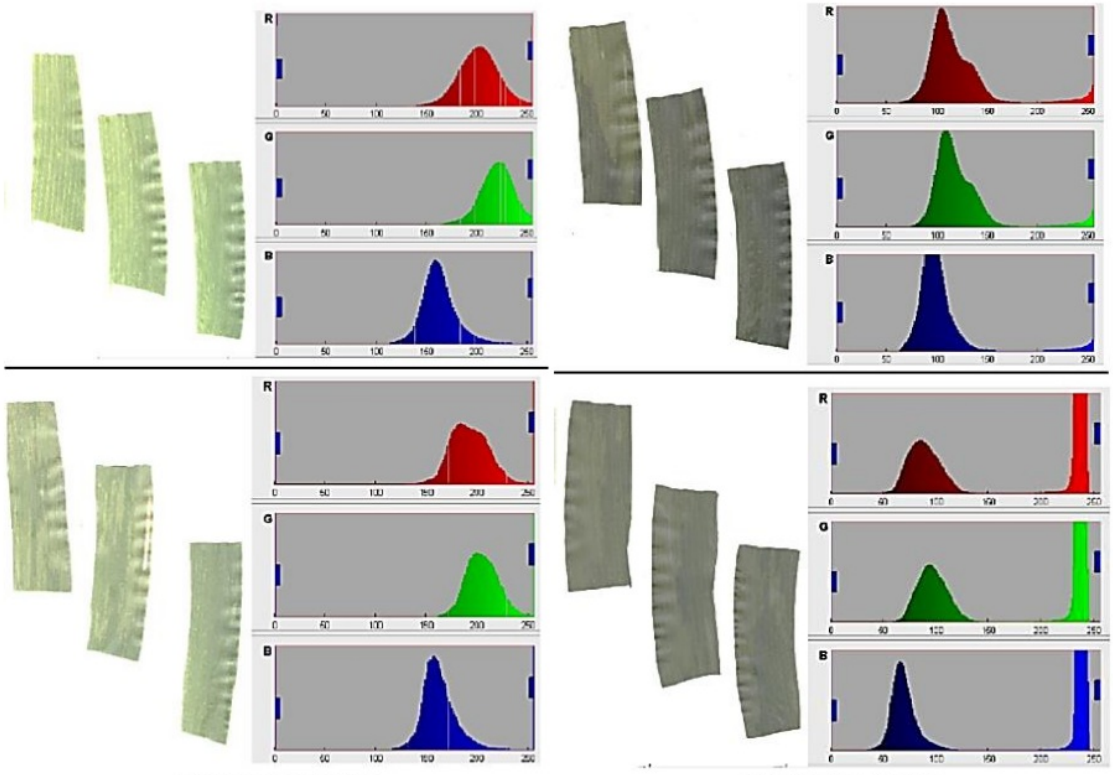
(a) Low (b) High

Figure 3. Concentration level of nitrogen

3. RESULTS AND ANALYSIS

In this section, the pre-processed leaf images were closely examined and multiple histograms of RGB scales are generated to observe similarity. Two sets of histograms corresponding to low and high concentration of nitrogen are shown in Figure 4.

Table 1 shows information (ranges) extracted from these histograms.

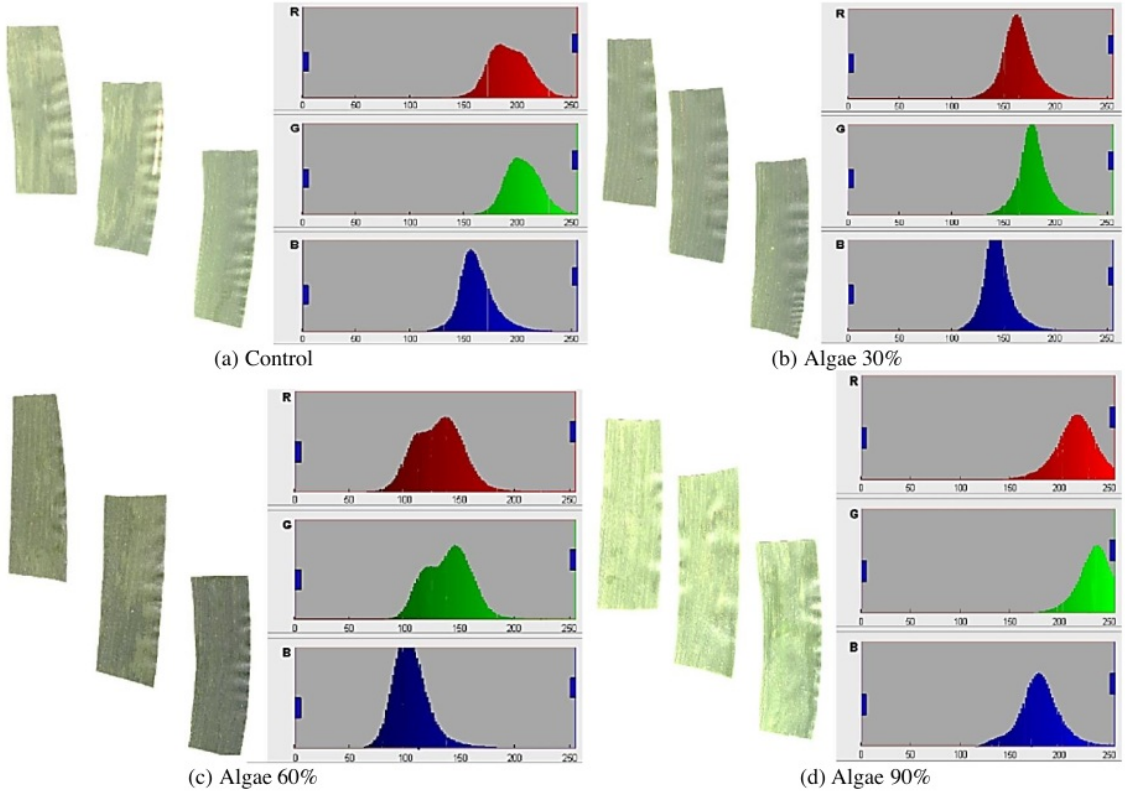


(a) Low concentration (b) High concentration

Figure 4. Two level of nitrogen concentrations based on corn leaf colour

Table 1. Level of nitrogen concentration level based on RGB

Colour / concentration	Low	High
Red	140-250	1st spike: 50-130, 2nd spike 200-250
Green	100-230	1st spike: 40-130, 2nd spike: 200-250
Blue	150-250	1st spike: 50-140, 2nd spike 200-250

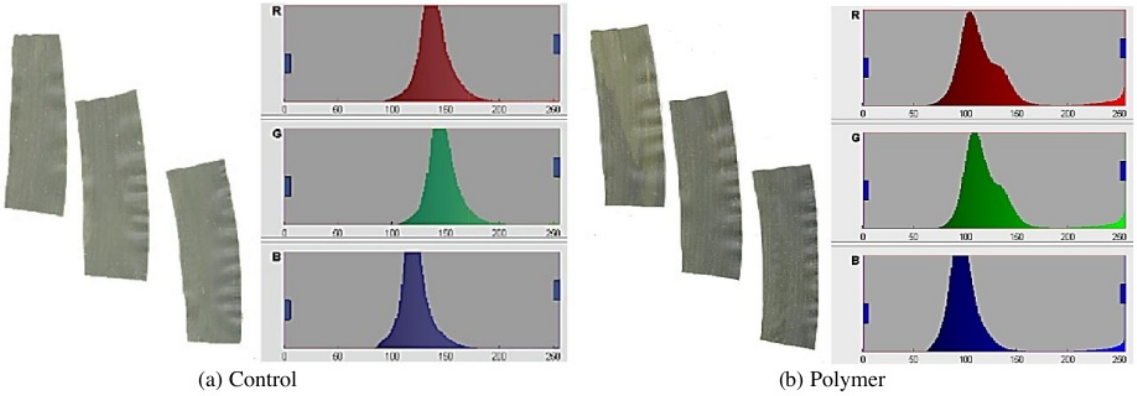


(a) Control (b) Algae 30%

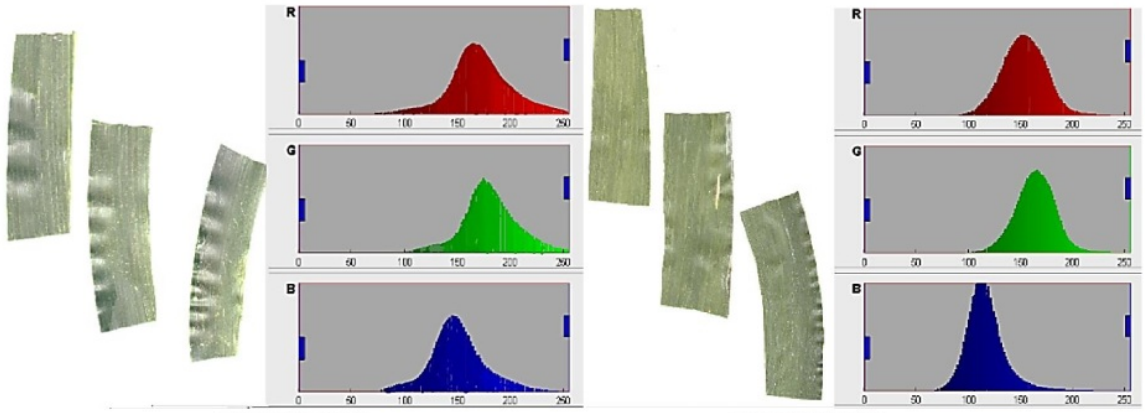
(c) Algae 60% (d) Algae 90%

Figure 5. Low level nitrogen concentration of corn fertilized by extract of algae

Two sets of experiments were performed, the first experiment involved manual comparison of the leaf with the green chart to identify the nitrogen level of a plant. The samples were grouped into two sets i.e. high and low nitrogen. The results were saved along with samples. The same experiment was performed using scanned image. The nitrogen level was categorized as high or low based on the given algorithm. After the results as shown in Figure 5, another data set was created to store all the generated results. These results were compared with the manual reading. It was found that computation gave the same results as the green chart up to 95%. There were more than one thousand samples that were used in the experiment before reaching a final accuracy rate. The green chart was then increased into the algorithm for better understanding and categorization of the result as shown in Figure 6.



(a) Control (b) Polymer

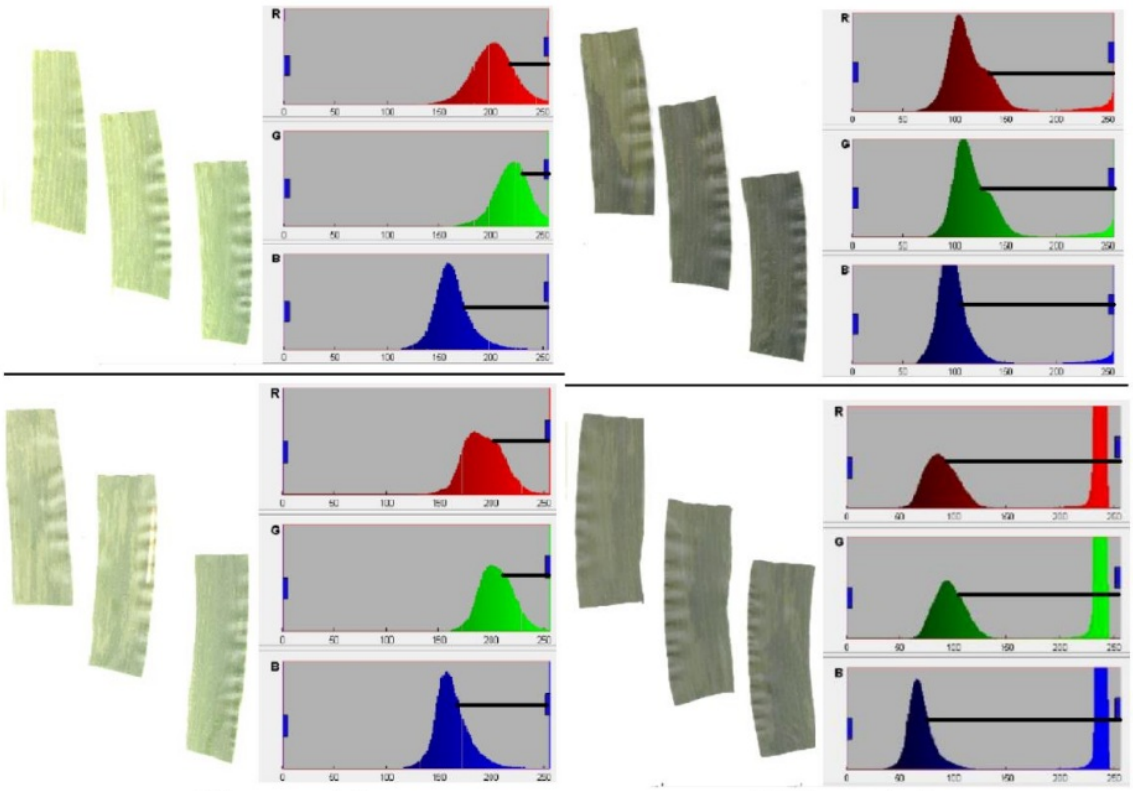


(c) Polymer 15% (d) Polymer 30%

Figure 6. High level nitrogen concentration of corn fertilized by polymer

The function of nitrogen in general is to stimulate plant growth and give the leaf a green color. The darker the green color on the leaves the more is the elemental nitrogen absorbed by the plant. In cereal plants, high nitrogen uptake can increase crop yields [20].

By using computer vision techniques, differences in the greenness of leaf color can be detected which are divided into 2 levels, namely low and high. This nitrogen level is based on a combination of Red, Green and Blue colors that form a bright or dark green color. Low levels of nitrogen found in corn fertilized with algae and darker leaf colors or higher nitrogen levels are found in corn fertilized with polymer. This indicates that the algae fertilized corn does not contain large amounts of nitrogen to be absorbed by plants. Instead, corn fertilized with polymers contains enough nitrogen to be absorbed by plants as seen in darker green leaves. The same thing is experienced by plants fertilized with NPK.



(a) Low concentration (b) High concentration

Figure 7. Difference between the graph peaks for low and high level concentration of nitrogen

As a comparison, nitrogen measurements using the Kjeldahl technique were also carried out in this study. The results in Table 2 shows that there are differences in measurement results for the corn fertilized with algae and polymer. The corn fertilized using polymers shows higher nitrogen levels when compared to corn fertilized with algae. In addition, the correlation of both measurement using Pearson correlation coefficient formula was used and it shows that there is strong positive correlation of both method which is shown by numbers 0.973 for low level concentration and 0.999 for high level concentration as shown on Table 3. This reveals that both measurements are correlated.

Table 2. Measurement nitrogen level of corn fertilized by algae and polymer fertilizer using Kjeldahl technique

Nitrogen concentration (%)			
Control	Algae (30%)	Algae (60%)	Algae (90%)
1.67	1.56	1.66	1.9
Control	Polymer	Polymer (15%)	Polymer (30%)
1.67	1.99	2.12	2.27

Table 3. Pearson correlation coefficients between image processing technique and Kjeldahl technique

Level/Colour	Red	Green	Blue
Low	0.973	0.973	0.973
High	0.999	0.999	0.999

Figure 4 shows the graphical representation of differences between high and low level of nitrogen. However, Figure 7 represents the difference between these graphs by using a backline. As shown in the graph, the spikes start from 50-80 rang. While that of low level starts after 100. The nitrogen level when low, the spikes starts at a higher value and when the nitrogen is high the spikes start at a smaller value.

4. CONCLUSIONS

Determining nitrogen level of a plant was a time inefficient process and not cost effective. This paper presents a time effecting and cost effecting technique for detection of nitrogen content in corn leaves by image procession and machine vison techniques, which been carried out by employing colour intensity levels in corn leaves. The technique used in this study uses existing leaf references to determine low or high nitrogen content. Leaf images obtained are processed to obtain the colour intensity of R, G and B in the leaves. The results of this colour intensity are compared with existing standard reference colours. Results show that the developed technique can accurately detect low or high nitrogen levels in corn leaves. Low nitrogen levels are indicated by leaves that are given algae fertilizer and high nitrogen levels are indicated by corn fertilized with NPK and polymer.

Currently how this technique could be deployed having limited computation power and in real time, acquiring similar result is in progress.

ACKNOWLEDGMENT

This work is supported by the Director General of Scientific and Technological Development and Advancement, Ministry of National Education, the Republic of Indonesia, batch No. 042.06.1.401516/2019.

REFERENCES

[1] Stuart, D., Schewe, R.L., McDermott, M. (2014). Reducing nitrogen fertilizer application as a climate change mitigation strategy: Understanding farmer decision-making and potential barriers to change in the US. Land Use Policy, 36: 210-218. https://doi.org/10.1016/j.landusepol.2013.08.011

[2] Fagodiya, R.K., Pathak, H., Kumar, A., Bhatia, A., Jain, N. (2017). Global temperature change potential of nitrogen use in agriculture: A 50-year assessment. Scientific Reports, 7: 44928. https://dx.doi.org/10.1038/srep44928

[3] Ma, L., Fang, J.L., Chen, Y.H., Gong, S.P. (2010). Color analysis of leaf images of deficiencies and excess nitrogen content in soybean leaves. In 2010 International Conference on E-Product E-Service and E-Entertainment, Henan, China, pp. 1-3. https://doi.org/10.1109/ICEEE.2010.5661235

[4] Lee, K.J., Lee, B.W. (2013). Estimation of rice growth and nitrogen nutrition status using color digital camera image analysis. European Journal of Agronomy, 48: 57-65. https://doi.org/10.1016/j.eja.2013.02.011

[5] Wang, Y., Wang, D.J., Zhang, G., Wang, J. (2013). Estimating nitrogen status of rice using the image segmentation of G-R thresholding method. Field Crops Research, 149: 33-39. https://doi.org/10.1016/j.fcr.2013.04.007

[6] Sun, Y.Y., Zhu, S.C., Yang, X., Weston, M.V., Wang, K., Shen, Z.Q., Xu, H.W., Chen, L.S. (2018). Nitrogen diagnosis based on dynamic characteristics of rice leaf image. PLOS ONE, 13(4). https://doi.org/10.1371/journal.pone.0196298

[7] Xu, G., Zhang, F.L., Shah, S.G., Ye, Y.Q., Mao, H.P. (2011). Use of leaf color images to identify nitrogen and potassium deficient tomatoes. Pattern Recognition Letters, 32(11): 1584-1590. https://doi.org/10.1016/j.patrec.2011.04.020

[8] Rorie, R.L., Purcell, L.C., Karcher, D.E., King, C.A. (2011). The Assessment of leaf nitrogen in corn from digital images. Crop Science, 51(5): 2174-2180. https://doi.org/10.2135/cropsci2010.12.0699

[9] Ulissi, V., Antonucci, F., Benincasa, P., Farneselli, M., Tosti, G., Guiducci, M., Tei, F., Costa, C., Pallottino, F., Pari, L., Menesatti, P. (2011). Nitrogen concentration estimation in tomato leaves by VIS-NIR non-destructive spectroscopy. Sensors (Basel, Switzerland), 11(6): 6411-6424. https://doi.org/10.2135/cropsci2010.12.0699

[10] Farneselli, M., Benincasa, P., Tei, F. (2010). Validation of N nutrional status tools for processing tomato. Acta Horticulturae, 852: 227-232. https://doi.org/10.17660/ActaHortic.2010.852.27

[11] Singandhupe, R.B., Rao, G.G.S.N., Patil, N.G., Brahmanand, P.S. (2003). Fertigation studies and irrigation scheduling in drip irrigation system in tomato crop (Lycopersicon esculentum L.). European Journal of Agronomy, 19(2): 327-340. https://doi.org/10.1016/S1161-0301(02)00077-1

[12] Padasali, V.A., Sunagar, V.B., Kattimani, P.A., Hiremath, N.V. (2014). Estimation of nitrogen content in

leaves using image processing. In Proceedings of International Conference on Advances in Engineering & Technology, Goa, India, pp. 25-28.
[13] Lemaire, G., Jeuffroy, M.H., Gastal, F. (2008). Diagnosis tool for plant and crop N status in vegetative stage: Theory and practices for crop N management. European Journal of Agronomy, 28(4): 614-624. https://doi.org/10.1016/j.eja.2008.01.005
[14] Sandoval-Villa, M., Guertal, E.A., Wood, C.W. (2000). Tomato leaf chlorophyll meter readings as affected by variety, nitrogen form, and nighttime nutrient solution strength. Journal of Plant Nutrition, 23(5): 649-661. https://doi.org/10.1080/01904160009382047
[15] Gianquinto, G., Sambo, P., Borsato, D. (2006). Determination of SPAD threshold values for the optimisation of nitrogen supply in processing tomato. Acta Horticulturae, 700: 159-166. https://doi.org/10.17660/ActaHortic.2006.700.26
[16] Han, S., Hendrickson, L.L., Ni, B. (2002). Comparison of satellite and artial imagery for detecting leaf chlorophyl content in corn. Transactions of the ASAE, 45(4). https://doi.org/10.13031/2013.9932
[17] Pettorelli, N. (2013). The Normalized Difference Vegetation Index. Oxford University Press.
[18] Sudhalakshmi, C., Velu, V., Thiyagarajan, T.M. (2008). Leaf colour chart for nitrogen management in rice - a review. Agricultural Review, 29(4): 306-310.
[19] Jumadi, O., Hala, Y., Iriany, R.N., Makkulawu, A.T., Baba, J., Inubushi, K. (2020). Combined effects of nitrification inhibitor and zeolite on greenhouse gas fluxes and corn growth. Environmental Science and Pollution Research, 27(2): 2087-2095. https://doi.org/10.1007/s11356-019-06776-6
[20] Mengel, K., Hutsch, B., Kane, Y. (2006). Nitrogen fertilizer application rates on cereal crops according to available mineral and organic soil nitrogen. European Journal of Agronomy, 24(4): 343-348. https://doi.org/10.1016/j.eja.2005.12.001

NOMENCLATURE

c	feature extraction variable
K	centre of the pixel
$S_{x,y}$	set of coordinates defining neighborhood $K(x,y)$
t	time

paper 1

ORIGINALITY REPORT

8%	5%	5%	1%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

MATCH ALL SOURCES (ONLY SELECTED SOURCE PRINTED)

2%

★ www.researchgate.net

Internet Source

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	On		

paper 1

GRADEMARK REPORT

FINAL GRADE

/0

GENERAL COMMENTS

Instructor

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

paper 2

by Yasser Djawad

Submission date: 24-Feb-2021 12:12PM (UTC+0500)
Submission ID: 1516869851
File name: SMT20120009_cp_1.pdf (1.03M)
Word count: 4498
Character count: 23449

www.ietdl.org

Published in IET Science, Measurement and Technology
Received on 24th January 2012
Revised on 13th April 2012
doi: 10.1049/iet-smt.2012.0009

ISSN 1751-8822

Robust feature extraction from impedimetric signals using wavelet packet decomposition with application to cytotoxicity testing

Y. Abdul Djawad J. Kiely M. Nibouche P. Wraith R. Luxton

University of the West of England, Coldharbour Lane, Bristol BS16 1QY, UK
E-mail: yasser.abddjawad@uwe.ac.uk

Abstract: Impedance-based cytotoxicity testing provides a simple, real-time and non-invasive assay technique that interferes minimally with cell morphology and function. In this study, impedimetric measurements of ECV304 cells with increasing concentrations of hydrogen peroxide were performed over 3 h, using D sensor connected to a lock-in amplifier at frequencies between 200 and 830 kHz. The output signals were analysed and decomposed using a two-level wavelet packet decomposition to obtain low- and higher-frequency coefficients of each signal. Subsequently, the sum of the negative and positive values and the difference values of each sub-band were calculated to acquire features of the signal coefficients. The patterns of features were compared with the microscopy images showing morphological changes of the cells for the four hydrogen peroxide concentrations and correlations between the extracted features and the morphological changes were identified. This demonstrated that wavelet packet decomposition was an effective feature extraction technique, which could provide a robust input vector for toxin clustering and classification purposes.

1 Introduction

The study of cell behaviour is important for many biological and biomedical applications, for instance to discover new drugs, to analyse cancer cells and for environment assessment. Numerous label-free techniques have been developed by researchers to study cell behaviour, including the use of micro- and nano-sensors [1], electric field imaging [2], microwave analysis [3] and impedance spectroscopy [4]. Among these techniques, impedance spectroscopy has been recognised as a powerful real-time technique to monitor cell behaviour. In addition, it is suitable for miniaturisation to meet the growing needs for micro-devices [5].

Impedance spectroscopy is a measurement technique for investigating the electrical properties of a material using electrically conducting electrodes [6]. Impedance spectroscopy has been applied in many biological research studies such as to analyse water for polluting parasites [7], to quantify cythopathic effects on Madin–Darby Carine Kidney (MDCK) cells [8], to investigate the human erythrocyte membrane [9], to detect oral cancer cells [5] and for cytotoxity testing [10, 11].

Impedance spectroscopy normally requires two electrodes: a counter electrode (CE) and a detecting electrode (DE). The basic methodology is to apply signal excitation via the CE and to observe the response on the DE. The impedance of the material can be measured over wide-range frequencies to study the characteristics of the material [12]. Using two electrodes technique, the impedance of electrodes has an influence on the measurement result. The measured impedance is the sum of the impedance of electrodes and the sample. To obtain a higher precision result, the four electrodes technique is developed [13]. However in this study, a high-frequency is required to penetrate the coverglass, therefore two electrodes technique is preferred since at high-frequency it generates less parasitic capacitance compared to four electrodes technique [14].

Various types of instruments have been used by researchers to perform impedance measurement such as LCR meters [15], impedance analysers [16] and lock-in amplifiers (LIA) [4, 12, 17, 18]. Impedance measurement systems based on LIA are very sensitive, can remove noise, minimise harmonic distortion and comprise low-cost apparatus [19]. LIA provide a measurement of an input signal which has the same frequency and phase as the reference signal.

Numerous techniques have been used to analyse biosignals. Wang *et al.* [20] used the short-time Fourier transform (STFT) to discriminate three types of cells (Jurkat, ECV304 and $CaCO_2$) and Jahankhani *et al.* [21] demonstrated the electroencephalogram signal feature extraction using a discrete wavelet transform (WT). Another approach was demonstrated by Tamil *et al.* [22] who applied the Hilbert Huang transform to perform signal feature extraction of an electrogastrogram.

For pattern recognition purposes, many methodologies are available. Most comprise of three processes: preprocessing, feature extraction and classification [23]. STFT has been shown to be an effective tool for signal feature extraction. It plots the time and frequency spectrum in a two-dimensional time–frequency plane representation. Nevertheless, STFT suffers from time and frequency resolution issues [24].

In this study, the WT is used to analyse the results of cytotoxicity testing experiments, since it is an appropriate technique to extract characteristics from time-varying biomedical signals [23] and it has capability to compress the parameters from a signal into a finite number of features [25]. In the present study, wavelet packet decomposition (WPD) has been applied to the data to gather low- and high-frequency components of the signals. The data were extracted into two levels using WPD and a statistical feature analysis was performed on each sub-band to obtain signal features for clustering and classification purposes.

2 Materials and methods

2.1 Cells preparation

ECV304 cells from the European Collection of Cell Cultures were seeded at density of 5×10^5 cells/ml in 75 cm^2 flask in 12 ml M199 media (Gibco) supplemented with 10% foetal calf serum and 2 ml L-glutamine and maintained in a humidified incubator at 37°C with 5% CO_2. The medium was changed every three days to feed the cells and monitored daily by microscope to check for confluence, at which point the cell monolayer covers 75–80% of flask surface. When confluence was achieved, the cell layer was washed with phosphate buffered saline followed by 1 ml of trypsin–ethylenediaminetetraacetic acid (EDTA) solution (0.05% porcine trypsin, 0.2 g/l EDTA) followed by incubation at 37°C for approximately 5 min to allow the cells to detach from the flask surface. Once the cells had detached, the trypsin was deactivated by the addition of 2 ml of growth medium and the cell density calculated by the trypan blue exclusion method. The cell suspension was diluted to a density of 5×10^5 cells/ml and 2 ml of cell suspension added to a LAB-TEK chamber and maintained at 35.7°C and 5% CO_2. Cytotoxicity testing was performed by adding dilute 10 mM hydrogen peroxide (H_2O_2) to the cell media with three concentrations 100, 500 and 1000 μM. The toxin was selected in order to evaluate cell injury that was caused by oxidative activity. Images were recorded every hour for 3 h using a Nikon phase contrast microscope at a magnification of 200. Image was used to measure cell area of ten cells in every image.

2.2 Instrumentation

The impedance measurements were conducted using an LIA connected to D-patterned sensor. The D pattern was selected as they provided greater response compared to interdigitated pattern because of the electric field strength distribution [11]. The D-patterned sensor was produced using PCB fabrication techniques and comprised a copper layer coated with 0.1 μm gold over 5 μm nickel. The diameter of each sensor was 22 mm, with 1 mm space between the CE and DE corresponding to the surface of the cell chamber. The D-patterned sensor and the chamber with the cells can be modelled as a capacitor as shown with yellow dotted line in Fig. 1, which the electrodes act as plates meanwhile the coverglass and the cells act as the dielectric. The D-patterned sensor has a capacitive value of about 2 pF. The series connection of the D-patterned sensor and a resistor creates an RC circuit. A 10 kΩ resistor was used to limit the current and to achieve a bandwidth about 8 MHz for the RC circuit, to ensure all measurement frequencies fell within this range. Measurement was performed by connecting the LIA in parallel with the sensors as illustrated in Fig. 1. The chamber was placed on the D sensor such that the cells were not in direct contact with the sensor, facilitating multiple, sequential measurement using a single sensor pair and to avoid cross contamination with the previous experiment. The LAB-TEK chamber base is made of borosilicate coverglass with thickness 0.13–0.16 mm and the culture area is 4.2 cm^2/well. It was critical that a thin coverglass was used to ensure that the electric field penetrated into the cell culture. During the experiments, the D sensor and the cell chamber were placed in the incubator to maintain their normal physiological state and metabolic function, preventing cell death.

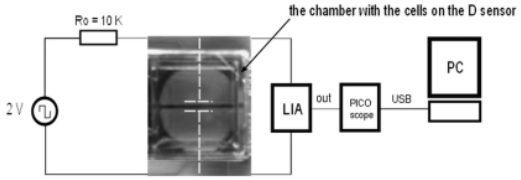


Fig. 1 *Circuit diagram of the system and a photograph of the D sensors below the cell chamber*

The LIA consists of filter, input amplifier, phase-sensitive detection and output amplifier. The high-pass filter (HPF) was set to have a frequency cutoff about 0.7 Hz to remove DC component of the input signal and the low-pass filter (LPF) was set to have a frequency cut off about 16 Hz to attenuate AC components and reject the noise on the LIA output. The output of LIA is a DC voltage. The input, a 2 V square wave signal, and the measurement of voltage output from the LIA were provided by PicoScope. PicoScope is a real-time PC-based digital oscilloscope using USB connection, which has capabilities to perform digital storage oscilloscope, metre and data logger, spectrum analyser and signal generator. The frequency was swept from 200 to 830 kHz at 10 kHz step intervals; in total there were 64 samples. This range was selected because beyond this frequency range, the results were similar for each H_2O_2 concentration or the outputs were inconsistent because of the effects of parasitic capacitances.

2.3 Wavelet packet feature extraction

WT is the technique used to transform a signal from the time domain into the time-scale/frequency plane. Continuous wavelet transform (CWT) compares the input signal by stretching and shifting the wavelet function to acquire the CWT coefficients. The wavelet extracts the information by obtaining the inner product of the signal with the wavelet function then shifting and repeating this process until the end of input signal. The process is then repeated from the beginning with a different scale to obtain different coefficients until the complete specified scale has been applied. However, calculating all CWT coefficients is highly redundant and requires a long computation time. Therefore discrete wavelet transform (DWT) is frequently used. In DWT, a series of HPFs and LPFs are used to separate high and low frequencies of the input signal. The result obtained from LPF is split again into high- and low-frequency components. Consequently, DWT decomposes low-frequency components only. To obtain more details about all coefficients including higher-frequency components, WPD is introduced. WPD decomposes the signal into approximation and detail equating to the low-

doi: 10.1049/iet-smt.2012.0009

frequency and high-frequency components, respectively. The process is repeated until each sub-band is decomposed into desired level of approximation and detail. Since the impedimetric data contain relevant information in the higher frequency, WPD is used. Two-level WPD extracts the signal into four sub-bands; low low (LL), low high (LH), high low (HL) and high high (HH).

A wavelet packet function is represented as follows [26]

$$\psi^i_{j,k}(t) = 2^{j/2}\psi^i(2^j t - k) \tag{1}$$

where i is the modulation parameter, j is the scale parameter and k is the translation parameter. The wavelet can be obtained by following repeated relationships

$$\psi^{2i-1}_{j+1}(t) = \sqrt{2}\sum_{k=-\infty}^{\infty} h(k)\psi^i_j(2t-k) \tag{2}$$

$$\psi^{2i}_{j+1}(t) = \sqrt{2}\sum_{k=-\infty}^{\infty} g(k)\psi^i_j(2t-k) \tag{3}$$

where $\psi^i(t)$ is the mother wavelet function, whereas $h(k)$ and $g(k)$ are quadrature mirror filters associated with scaling function and wavelet function [27]. The wavelet packet coefficient C of the signal $f(t)$ then is described as follows

$$C^i_{j,k} = \int_{-\infty}^{\infty} f(t)\psi^i_{j,k}(t)\,\mathrm{d}t \tag{4}$$

3 Results and discussion

Following treatment with H_2O_2, impedance measurement and images were recorded every hour for 3 h. Result from cells exposed to 0 μM (control), 100, 500 and 1000 μM H_2O_2 were compared to determine the effect of H_2O_2 on the cells.

3.1 Image analysis

Images were captured and cell areas were calculated for ten cells in each image at 1, 2 and 3 h for each concentration of H_2O_2. The images, shown in Fig. 2, indicate increasing cell disruption with increasing concentrations of H_2O_2 from 100 to 500 μM. Interestingly, less disruption was noted at 1000 μM H_2O_2. Cell disruption was characterised by shrinkage at the cell membrane and increased 'blebbing' of the cell membrane, resulting from the cytotoxic effect of the H_2O_2. When the cells were exposed to 1000 μM H_2O_2, the cells displayed less morphological change as the high level

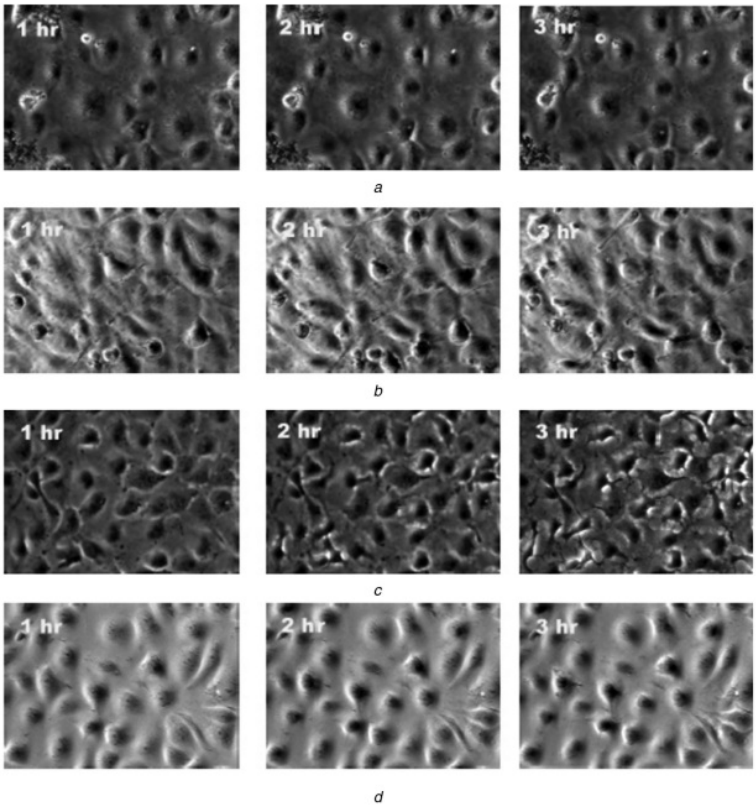


Fig. 2 *Images of ECV304 cells with increasing concentrations of H_2O_2 for 3 h at intervals of 1 h*

a 0 μM
b 100 μM
c 500 μM
d 1000 μM

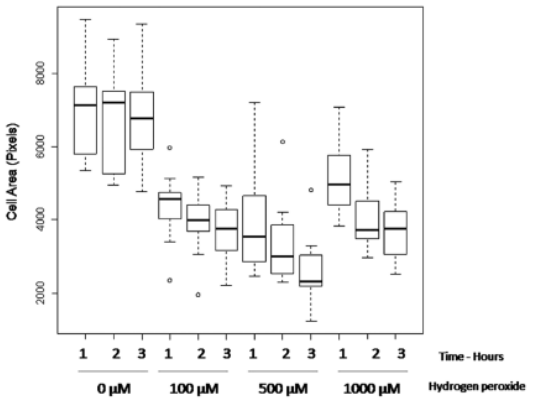


Fig. 3 *Box and Whiskers plot of cell size measurement with different H_2O_2 concentrations*

of toxin-prevented cell mechanisms occurring that result in cell shrinkage and 'bleb' formation. Fig. 3 shows Box and Whiskers plot of the cell area derived from the images shown in Fig. 2.

The result of image analysis shows that there was a significant reduction in cell size for all concentrations of H_2O_2 compared with the control (0 µM H_2O_2). Significant reductions in cell area were also seen over time. With 100 and 500 µM H_2O_2 there were a significant reduction in cell area over the 3 h. In comparison, there was not significant change in cell area over 3 h in the control. At the highest concentration of H_2O_2, there was a significant reduction in cell size between 1 and 2 h, but then stabilises by the third hour showing that the cell processes have been inhibited after 2 h.

3.2 Wavelet packet features

Fig. 4 shows the voltage output of the LIA for the cells with added H_2O_2 at concentrations of 0, 100, 500 and 1000 µM. The plots show a resonant frequency around 400 kHz. Below the resonant frequency, the output voltage decreases linearly from about 1000 to −1000/−1500mV. At this point, there is a clear differentiation between the signals at different times and toxin concentrations. Beyond the resonant frequency, the voltage output gradually rises to approximately 500 mV at around 520 kHz. At this frequency, the amplitude tends to decline until frequency 700 kHz, where again the differentiation can be seen clearly. After this stage, the output amplitude rises monotonically again until a frequency of 830 kHz.

The measurement scans were captured for each data set every hour for 3 h after H_2O_2 was added. Each signal was sampled 64 times. Instead of 16 samples at the final stage, the output of each band at level 2 has 27 samples because of convolution process between signal and filter function. In the present study, the signal is extracted into two levels. Two-level decomposition was chosen since a higher level provides almost identical results. The signals were analysed using various wavelet functions and sym8 [28] was selected for the reason that it preserves signal features like the amplitudes, which is important in this application.

To reduce the number of vectors for clustering and classification purposes, statistical analysis was performed

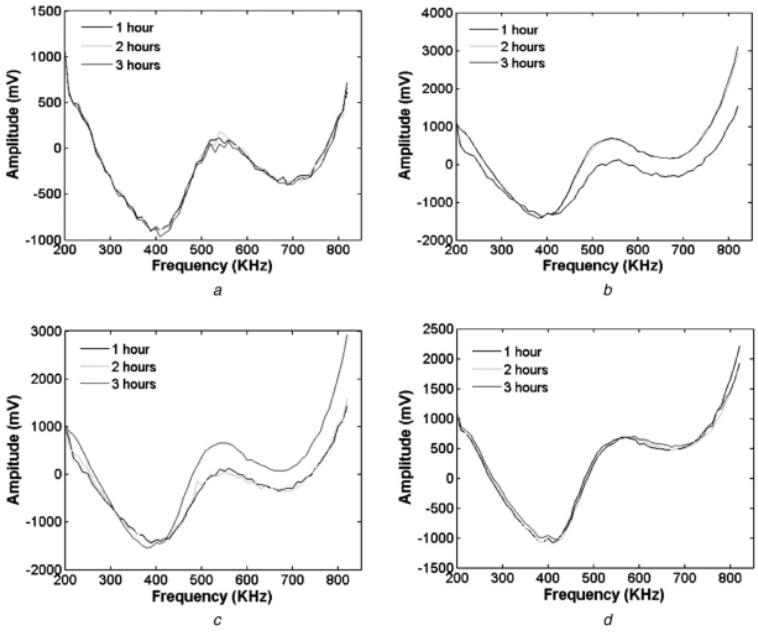


Fig. 4 *Output voltage of LIA across D sensor and ECV304 cells with four concentrations of H_2O_2*

a 0 µM
b 100 µM
c 500 µM
d 1000 µM

doi: 10.1049/iet-smt.2012.0009

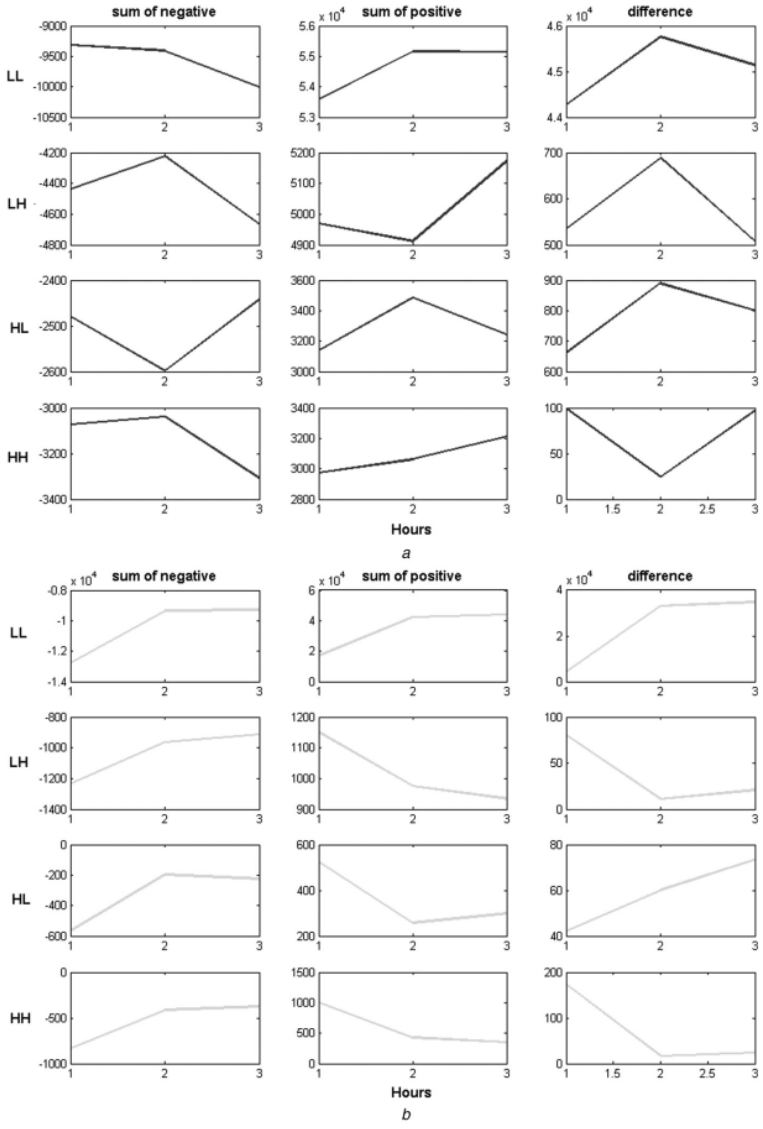


Fig. 5 *Combined statistical features of ECV304 cells with four different concentrations of H_2O_2 for 3 h*

a 0 μM
b 100 μM
c 500 μM
d 1000 μM

on the WPD results to obtain time–frequency distributions. In order to analyse how the data was distributed on the vertical axes the following statistical features were computed:

- sum of negative values of each sub-band;
- sum of positive values of each sub-band;
- difference between negative and positive values of each sub-band.

This resulted in 36 statistical features for each H_2O_2 concentration, as shown in Fig. 5. Seven types of pattern of change in the slope with respect to time are evident. These are distributed uniquely for each concentration. The cells + 0 μM H_2O_2 exhibit five types of pattern, three types of pattern were in sum of negative features, four types of pattern were in the sum of positive features and two types of pattern were in the difference, with three types common patterns. The cells + 100 μM H_2O_2 had three types of pattern, one type of pattern was in the sum of negative features, two types of patterns were in sum of positive features and three types of pattern were in difference with two types common patterns. The cells + 500 μM H_2O_2 had four types of pattern, two types of pattern were in the sum of negative features, three types of pattern were in the sum of positive features and four types of pattern were in the difference with three types common patterns. Finally, cells + 1000 μM H_2O_2 had five types of pattern, three types

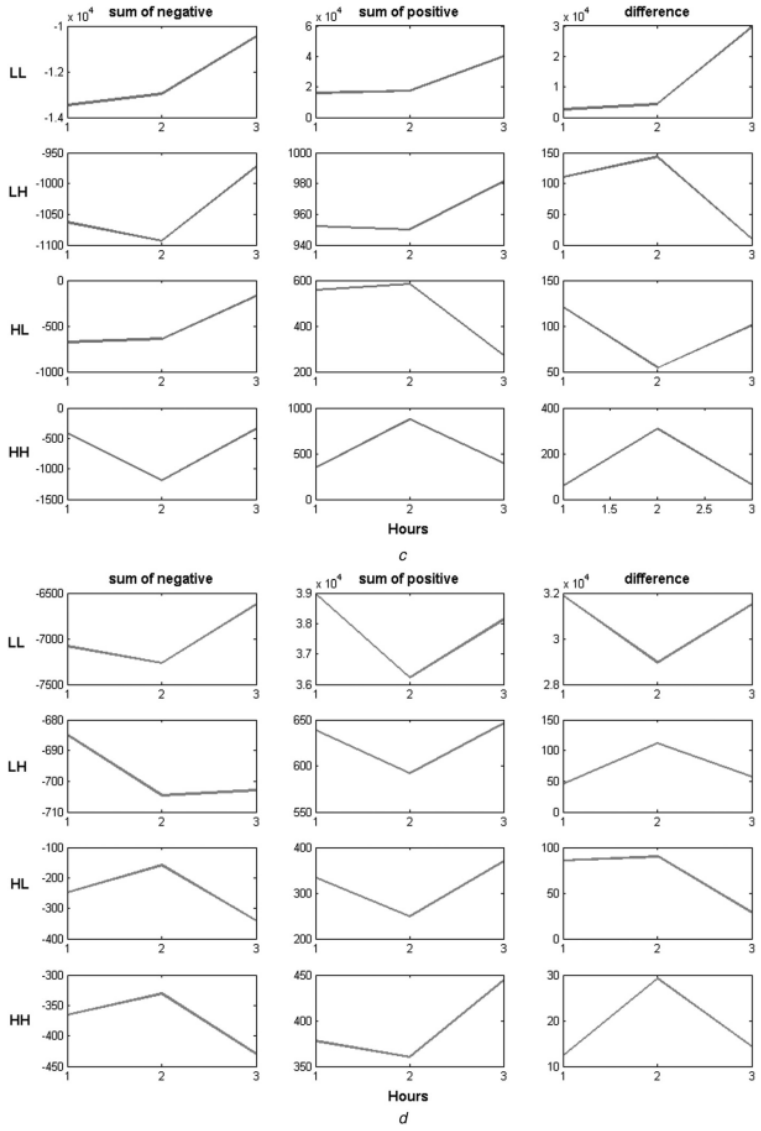


Fig. 5 *Continued*

of pattern were in the sum of negative features, two types of pattern were in the sum of positive features and three types of pattern in the difference with three types common patterns.

More than half of the cells + 0 μM H_2O_2 patterns shown in Fig. 5*a* demonstrated a small change from hour to hour compared with the higher concentrations. The patterns for the cells + 100 μM H_2O_2 in Fig. 5*b* demonstrated either increasing or decreasing values from the first to the second hour and a small (approximately constant) value change from the second to the third hour. There is a single exception for HL in the difference column. The cells + 500 μM H_2O_2 produced the opposite result. Most of the patterns do not change significantly from the first to the second hour, whereas change is evident from the second to the third hour. Only four patterns showed considerable change over 3 h period. The results for the cells + 1000 μM H_2O_2 show that most of the patterns changed over small range each hour. The experiment has been done twice and provided considerably similar result.

3.3 Correlation between cell size and WPD statistical features

Correlation coefficients between cells size measured from microscopy images and wavelet packet features are shown in Table 1. The Pearson's correlation coefficient has been used in this study and defined as follows

$$r_{xy} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - \left(\sum x_i\right)^2}\sqrt{n\sum y_i^2 - \left(\sum y_i\right)^2}} \quad (5)$$

doi: 10.1049/iet-smt.2012.0009

Table 1 Correlation coefficients between cell size and wavelet packet features

Features	LL	LH	HL	HH
sum of positive 1 h	0.9696	0.8834	0.8994	0.8961
sum of positive 2 h	0.8619	0.9682	0.9587	0.8931
sum of positive 3 h	0.9369	0.9400	0.9847	0.9560
sum of negative 1 h	−0.6070	0.8887	0.8607	0.8901
sum of negative 2 h	−0.2820	0.9608	0.9369	0.9293
sum of negative 3 h	0.1371	0.9408	0.9696	0.9481
difference 1 h	0.9438	0.7510	0.9758	−0.5708
difference 2 h	0.8238	0.9840	0.9939	0.6580
difference 3 h	0.9948	0.9327	0.9727	−0.9798

where x and y are the data values and n is number of data points. The results range between -1 and $+1$; a result close to -1 or $+1$ indicates a strong correlation. The data in Table 1 demonstrate the relationship between the cell size measurement from Fig. 3 and the wavelet packet features in Fig. 5.

In general, the table shows strong correlation between cell size and wavelet packet features indicating that the feature extraction technique using wavelet packet is an effective method of predicting cell size. The closest relationships are provided by the sum of positive features indicated by the fact that in all the wavelet sub-bands, the correlation values are close to $+1$. Among the four sub-bands, features in sub-band HL show the strongest correlation. Weak correlations are found in the sum of negative values of sub-band LL and the difference of sub-band HH. It is also shown that at 3 h, nearly all the features demonstrated high correlation coefficients greater than 0.9 in each sub-band. This could indicate that at 3 h the cells have reached a steady state. Consequently, cell size are consistent and size measurement from the images correlates closely with wavelet packet features.

4 Conclusions

An impedimetric measurement has been performed over 3 h on ECV304 cells with four different concentrations of H_2O_2 (0, 100, 500 and 1000 μM). Time frequency representations of the non-stationary signals have been extracted using the two-level WPD. Subsequently, the signals were statistically analysed to establish key features. The technique revealed that each sub-band has a unique pattern of features pattern, depending on concentration of H_2O_2. The results of cell size measurement on microscopy images were compared with the statistical features. Nearly all the statistical features showed good correlation with the cell size; the HL sub-band had high particularly correlation coefficients. Consequently, the WPD technique provides an accurate non-visual method of monitoring changes in cell morphology following a toxin challenge. This indicate that the proposed technique promises to be an effective tool for clustering and classification purposes. A further study with different types of toxin is underway.

5 Acknowledgments

Y.A. Djawad thanks the Indonesian Higher Education Department for the funding support during his study.

6 References

1 Senez, V., Arscott, S.: 'Evaluation of electric impedance spectra for single bio-cells in microfluidic devices using combined FEMLAB/ELDO modeling'. Proc. COMSOL Multiphysics User's Conf., Paris, France, 2005

2 Ghallab, Y.H., Badawi, W.: 'Techniques for biocells sensing, detection and characterization'. Proc. 2005 Int. Conf. on MEMS, NANO and Smart System (ICMEMNS 2005), 2005, pp. 402–405

3 Dalmay, C., Cheray, M., Pothier, A., Lalloue, F., Jauberteau, M.O., Blondy, P.: 'Ultra sensitive biosensor based on impedance spectroscopy at microwave frequencies for cell scale analysis', *Sens. Actuators A, Phys.*, 2010, **162**, pp. 189–197

4 Giaever, I., Keese, C.R.: 'Micromotion of mammalian cells measured electrically'. Proc. National Academy of Sciences of the United States of America 1991, vol. 88, pp. 7896–7900

5 Arias, L.R., Perry, C.A., Yang, L.: 'Real time electrical impedance detection of cellular activities of oral cancer cells', *Biosens. Bioelectron.*, 2010, **25**, pp. 2225–2231

6 Macdonald, J.R.: 'Impedance spectroscopy', *Ann. Biomed. Eng.*, 1992, **20**, pp. 289–305

7 Houssin, T., Follet, J., Follet, A., Dei-Cas, E., Senez, V.: 'Label-free analysis of water-polluting parasite by electrochemical impedance spectroscopy', *Biosens. Bioelectron.*, 2010, **25**, pp. 1122–1129

8 McCoy, M., Wang, E.: 'Use of electrical cell-substrate impedance sensing as a tool for quantifying cytophatic effect in influenza a virus infected MDCK cells in real-time', *J. Virulogic. Methods*, 2005, **130**, pp. 157–161

9 Ivanov, I.T.: 'Impedance spectroscopy of human erythrocyte membrane: effect of frequency at the spectrin denaturation transition temperature', *Bioelectrochemistry*, 2010, **78**, pp. 181–185

10 Xiao, C., Lachance, B., Sunaharta, G., Luong, J.H.: 'Assessment of citotoxicity using cell-substrate impedance sensing:concentration and time response function approach', *Anal. Chem.*, 2002, **74**, pp. 5748–5753

11 Vargas, M.A.M.: 'Real time monitoring of cellular toxicity using electrical impedance measurement'. PhD thesis, University of the West of England, Bristol, UK, 2007

12 Luong, J.H.T.: 'An emerging impedance sensor based on cell-protein interactions: application in cell biology and analytical biochemistry', *Anal. Lett.*, 2003, **36**, pp. 3147–3164

13 Laugere, F., Lubking, G.W., Bastemeijer, J., Vellekoop, M.J.: 'Design of an electronic interface for capacitively coupled four-electrode conductivity detection in capillary electrophoresis microchip', *Sens. Actuators B*, 2002, **83**, pp. 104–108

14 Pliquett, U., Nacke, T., Frense, D., *et al.*: 'Evaluation of electrodes for impedance spectroscopy', *IFMBE Proc.*, 2009, vol, **25**, pp. 251–254

15 Bauchot, A.D., Harker, F.R., Arnold, W.M.: 'The use of electrical impedance spectroscopy to assess the physiological condition of kiwifruit', *Postharvest Biol. Technol.*, 2000, **18**, pp. 9–18

16 Zhang, T., Wang, R.Y., Bao, Q.Y., Rawson, D.M.: 'Development of new rapid measurement technique for fish embryo membrane permeability studies using impedance spectroscopy', *Theriogenology*, 2006, **66**, pp. 982–988

17 Arndt, S., Seebach, J., Psathaki, K., Galla, H., Wegener, J.: 'Bioelectrical impedance assay to monitor changes in cell shape during apoptosis', *Biosens. Bioelectron.*, 2004, **19**, pp. 583–594

18 Xiao, C., Luong, J.H.T.: 'A simple mathematical model for electric cell-substrate impedance sensing with extended applications', *Biosens. Bioelectron.*, 2010, **25**, pp. 1774–1780

19 Prodromidis, M.I.: 'Impedimetric immunosensors a review', *Electrochim. Acta*, 2010, **55**, pp. 4227–4233

20 Wang, Z., Kiely, J., Nibouche, M., Luxton, R.: 'Impedimetric discrimination of cell types for use in a whole cell biosensor'. Proc. Tenth World Congress on Biosensors, Shanghai, China, 2008

21 Jahankhani, P., Kodogiannis, V., Revett, K.: 'EEG signal classification using wavelet feature extraction and neural network'. Proc. IEEE John Vincent Atanasoff 2006 Int. Symp. on Modern Computing (JVA'06), 2006, pp. 120–124

22 Tamil, E.M., Hamzah, R., Idris, M.Y.I., Tamil, A.M.: 'Feature extraction for biosignal processing (Part IV: electrogastrogram) using HHT'. Proc. Fourth Kuala Lumpur Int. Conf. on Biomedical Engineering, Kuala Lumpur, Malaysia, 2008, pp. 195–198

23 Cvetkovic, D., Ubeyli, E.D., Cosic, I.: 'Wavelet transform feature extraction from human PPG, ECG, and EEG signal responses to ELF PEMF exposures: a pilot study', *Digit. Signal Process.*, 2008, **18**, pp. 861–874

24 Narasimhana, S.V., Pavanalathab, S.: 'Estimation of evolutionary spectrum based on short time Fourier transform and modified group delay', *Signal Process.*, 2004, **84**, pp. 2139–2152

25 Daubechies, I.: 'The wavelet transform, time-frequency localization and signal analysis', *IEEE Trans. Inf. Technol.*, 1990, **36**, pp. 961–1005
26 Ogden, R.T.: 'Wavelet packets', in 'Essential wavelets for statistical applications and data analysis' (Birkhaeuser-Boston, 1997, 1st edn.), p. 174
27 Shen, M., Chen, J., Beadle, P.J.: 'Analysis of time-varying EEG based on wavelet packet entropy'. Proc. Sixth Int. Symp. on Neural Network, 2009, China, pp. 21–28
28 Daubechies, I.: 'Ten lectures on wavelets'. CBMS-NSF Regional Conf. on Ser. Applied Mathematics, 1992

doi: 10.1049/iet-smt.2012.0009

paper 2

ORIGINALITY REPORT

11%	7%	8%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

MATCH ALL SOURCES (ONLY SELECTED SOURCE PRINTED)

1%

★ Bae, K.U., Y.S. Won, and N.H. Myung. "Design method for bandpass filter with enhanced stopband rejection using spiral SIRs", Electronics Letters, 2012.

Publication

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	On		

paper 2

GRADEMARK REPORT

FINAL GRADE

/0

GENERAL COMMENTS

Instructor

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

PAGE 8

paper 3

by Yasser Djawad

Submission date: 24-Feb-2021 12:13PM (UTC+0500)
Submission ID: 1516870163
File name: Yasser_Djawad_IRBM.pdf (1.64M)
Word count: 4870
Character count: 24782

IRBM 42 (2021) 28–34

Contents lists available at ScienceDirect

IRBM

www.elsevier.com/locate/irbm

Elsevier Masson France

EM|consulte

www.em-consulte.com

Original Article

Development of an Intelligent Mobile Health Monitoring System for the Health Surveillance System in Indonesia

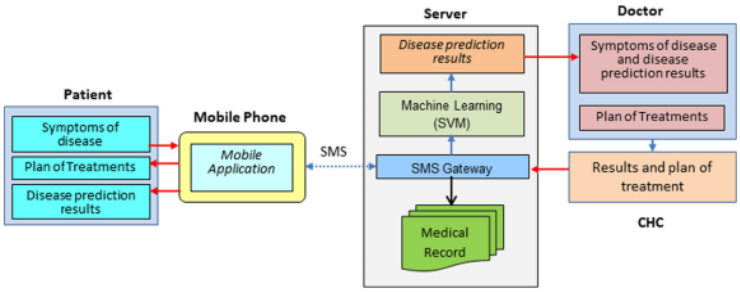
Y.A. Djawad [a,*], S. Suhaeb [a], Ridwansyah [a], H. Jaya [a], Fathahillah [b], Saharuddin [a]

[a] Department of Electronic Engineering, Universitas Negeri Makassar, Indonesia
[b] Department of Computer Engineering, Universitas Negeri Makassar, Indonesia

HIGHLIGHTS

- Facilitating health care in remote and rural area using mobile phone infrastructure.
- Prediction of disease based on data sends by patient using machine learning (Support Vector Machine).
- Characterise the proposed model with different type of kernels and parameters to find optimum classification.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 19 December 2018
Received in revised form 21 August 2020
Accepted 4 October 2020
Available online 9 October 2020

Keywords:
Telehealth
ICT
Health care
Kernels

ABSTRACT

Mobile phone applications have been widely used in various fields, including health care. Generally, this technology is used to overcome problems in health care by utilising mobile phone features for facilitating basic needs in health services. This study proposes an intelligent mobile health monitoring system that can be used in rural and remote areas where health services are still lacking. The system was made based on client/server architecture. Nine symptoms of typhoid, cough and diarrhoea from 30 patients were gathered from a hospital. Based on this data, a machine learning model using Support Vector Machine (SVM) was performed to distinguish these diseases. To find the best model parameters of the SVM, three different kernels (linear, polynomial, and Radial Basis Function (RBF)) were analysed. The result showed that RBF with degree 2 provided the best result in this particular application. The system was designed to receive input from patients about symptoms of the disease they have. The mobile phone application sends the data of the symptoms using Short Message Service (SMS) to the server. Furthermore, a machine algorithm module in the server identifies to which disease it belongs to based on the machine learning model created before. The prediction result is accessible to the doctor and the nearest Community Health Center (CHC). Based on the result, the doctor proposes a treatment plan for the patient to be recorded and sent to the patient by CHC. The proposed mobile health monitoring system has run properly and is ready to be evaluated in a real situation.

1. Introduction

Nowadays, health care has become humans basic need other than clothing, food, and housing because health is an important factor which supports individuals' daily activities. Therefore, the need for health care is inevitable. Health care is affected by various

* Corresponding author at Universitas Negeri Makassar, Jl. A.P. Pettarani, Makassar 90222, Indonesia.
E-mail address: yasser.djawad@unm.ac.id (Y.A. Djawad).

https://doi.org/10.1016/j.irbm.2020.10.001
1959-0318/

factors such as the availability and quality of medical instruments, health care workers and health care facilities. Along with the increasing number of the population that results in various complex problems, including health problems, the demand for health care rapidly increases [1]. However, more health care problems are faced by people living in remote and rural areas where health care facilities are poor [2,3]. Thus, innovation is required to provide an appropriate method to deliver health care to remote and rural areas.

The increasing use of mobile applications provides an opportunity to apply mobile applications in health care. Various techniques have been used by researchers to use mobile application since they consist of audio, video and data processing technology that can be utilised to improve health care. In Indonesia, some researchers have conducted studies to make telehealth system and mobile phone applications. An Integrated e-Prescription system has been created to improve health care for community health centres (CHCs) [4]. An m-Health system was created by [5] to support mother and child care at CHCs. To obtain data from health workers, a study was conducted to obtain nurses' perception of cardiovascular cases [6]. A monitoring system using a mobile application was created to support hypertension treatment [7]. The system would provide data to a doctor to be analysed. A smart telehealth system was developed utilising mobile phone application to perform fetal detection and biometry measurement using Adaptive Mahalonobis Generalized Learning Vector Quantization (AMGLVQ) [8]. The use of telemedicine was applied during natural disasters using several types of telecommunications networks such as radio, satellite, GSM, CDMA and the internet [9]. This device provides high accessibility to be used in various conditions even though this would cost much. This tool is also equipped with various medical devices such as ECG monitors, SpO2, fetal heart rate and blood pressure device. Therefore, this application requires someone who is trained to operate it. A store and forward method was applied in telemonitoring fetal growth in rural areas [10]. The store-forward method requires an internet connection that is not real-time.

However, all previous studies in Indonesia have not applied machine learning algorithms to automatically predict common diseases which almost have similar symptoms that many people suffer from such as typhoid, cough and diarrhoea. In addition, none of them has utilised Short Message Service (SMS) which does not require Internet technology. While mobile technology infrastructures have been widely installed in remote and rural areas in Indonesia. Even though, the latest technology for mobile phones such 4G has not been installed yet. Therefore, a proper technique is needed to utilise the installed mobile phone technology in remote and rural areas. Mobile text technology is a promising technology to use in these circumstances. SMS is the standard communication protocol used by mobile phones to exchange short text messages. This SMS technology is well known to mobile phone users because it is easy to use and inexpensive. The SMS technology can also be used on all types of 2G, 3G and 4G networks so that it can be used by users in remote and rural areas where mobile phone infrastructure already exists but cannot access an adequate internet network.

In this study, a classification algorithm is applied to predict the disease of the patients. Classification is needed when a set of disease features is hard to separate where they belong to. A classification model was created as a base for disease prediction. Various learning machine algorithms were used by the researchers for cancer and disease classification. One of the algorithms is the Support Vector Machine (SVM) [11]. SVM provides flexibility by introducing the kernel for separating data and by selecting an appropriate kernel, the similarity between data can be stressed to obtain a higher value of the kernel [12]. SVM has been used in the biomedical field. Two machine learning (ML) algorithms were compared; they are Random Forest (RF) and Support Vector Machine (SVM) for microarray-based cancer classification. The result showed that SVM performed better than RF in this particular application [13,14]. A similar result was found when applying ML in diabetes research, the results showed that SVM performed better than other ML algorithms in the study [15]. However, there is no single algorithm that is suitable for all applications.

2. Materials and methods

2.1. System design

The concept of the intelligent mobile health monitoring system is to create a simple mobile application to send the symptoms of a disease by patients and a system that able to predict the disease based on the symptoms sent. Furthermore, the doctors can access the result and give a plan of treatment. The CHC receives the result and plan of treatment to be processed and to be sent to the patient. The proposed mobile health monitoring system was divided into 2 sides; client-side and server side as shown in Fig. 1. The mobile application was the client-side that gathers symptoms data from patients and sends the data to the server using text message services. In the server-side, a program was created to classify the data received from the mobile application. The data sent by the patient were compared to the classification predictive model that had been created using the SVM algorithm to determine to which membership the data belong.

Patients symptoms or attributes such as temperature (TEMP), cough (SYMCOUGH), headache (SYMHA), fever (SYMFEV), chest pain (SYMCP), cough with blood (SYMCWB), limp (SYMLIMP), constipation (SYMCON) and stomach ache (SYMSMA) were obtained from Hospital of Hajj, Makassar, Indonesia. This record contains 9 symptoms which were collected from the records of 3 diseases; typhoid, cough and diarrhoea. The diseases were chosen as they were dominant during the data collection. There were 10 patient records for each disease. The data had been rigorously checked by the hospital local committee. The data were used to create a predictive model for classification purposes using the SVM algorithm.

2.2. Support Vector Machine (SVM)

The SVM is a supervised machine learning algorithm used for classification purposes. The essence of this technique is to separate the data with maximum margin on the boundary decision and to minimise the number of misclassified data samples.

It is assumed that there are L training points where each input $\mathbf{x}_i$ and it is written in the form

$$\{\mathbf{x}_i, y_i\}, i = 1 \ldots L \tag{1}$$

where object $\mathbf{x}_i$ is called *pattern* and y_i is the label associated with $\mathbf{x}_i$. It is assumed that the data are linearly separable, which means that there is a line that can be drawn between these data to separate them. A linear classifier is defined as the *dot product* between two vectors $\mathbf{w}$ and $\mathbf{x}$ as $\mathbf{w}^T\mathbf{x}$ and can be defined as

$$f(t) = \mathbf{w}^T\mathbf{x} + b \tag{2}$$

The vector $\mathbf{w}$ is the *weight vector* and b is the *bias*. A line that separates the two vectors or divides the space into two is known as *hyperplane* as shown in Fig. 2 and can be written as follows:

$$\{x : f(x) = \mathbf{w}^T\mathbf{x} + b = 0\} \tag{3}$$

The points which lie on the planes H_1 and H_2 are also known as the *support vectors* and the planes H_1 and H_2 are described as:

$$\begin{aligned} \mathbf{x}_i\mathbf{w} + b &= +1 \quad H_1 = +1 \\ \mathbf{x}_i\mathbf{w} + b &= -1 \quad H_2 = -1 \end{aligned} \tag{4}$$

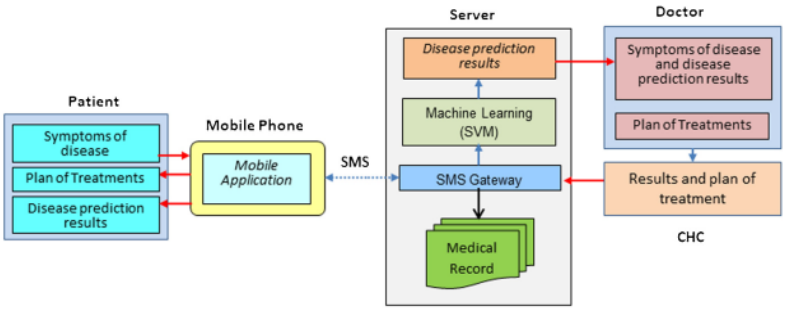


Fig. 1. System architecture of the intelligent mobile health monitoring system.

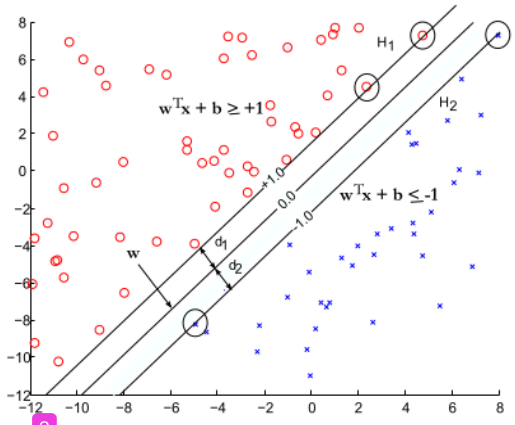


Fig. 2. The circled data points are the *support vectors* which are actually the closest points to the decision boundary.

The distance between H_1 and H_2 are defined as d_1 and d_2. These distances are known as the *margin* of SVM. These margins have to be as far as possible from the hyperplane to separate two vectors or classes. The maximum margin classifier is defined by maximising the margin $1/\|\mathbf{w}\|$ which is similar to minimising $\|\mathbf{w}\|^2$. This is also known as the hard margin and defined as

$$\underset{\mathbf{w},b}{\text{minimise}} \quad \frac{1}{2}\|w\|^2 \tag{5}$$

A problem arises when the data can not be separated using a linear classifier or hyperplane. A non linear classifier is introduced to overcome this problem. The data are distinguished by mapping the data from input space X to higher dimensional space F using non-linear mapping function ϕ and equation (2) is changed to:

$$f(t) = \mathbf{w}^T\phi(\mathbf{x}) + b \tag{6}$$

If weight vector $\mathbf{w}$ is expressed as a linear combination of training examples $\mathbf{w} = \sum_{i=1}^{n}\alpha_i\mathbf{x}_i$, then Equation (6) can be written in the form:

$$f(t) = \sum_{i=1}^{n}\alpha_i\phi(\mathbf{x}_i)^T\phi(\mathbf{x}_i) + b \tag{7}$$

where *kernel function* is defined as:

$$k(\mathbf{x}, \mathbf{x}_i) = \phi(\mathbf{x}_i)^T\phi(\mathbf{x}_i) \tag{8}$$

Some popular kernels are Gaussian, Polynomial, Sigmoidal and Radial Basis Function (RBF).

3. Results

The use case diagram of the proposed mobile health monitoring system is shown in Fig. 3. It shows the patient's role compared to others. The user diagram describes the relationship between use cases and actors of the proposed mobile health monitoring system. The first actor is a patient who should give information about the symptoms of the disease. The new patient should register to the system before accessing the service of the system. The new patient should apply to the system by filing a personal form including national health insurance id if he or she has it. The patient data is sent to the server using an SMS. The CHC would approve the application. The registered patient can access the system by filing the symptoms form regarding when he or she gets a disease. The symptoms data filled by the patient are sent to the server also using an SMS. The data would be processed by the server. The use of the server in this application aims as a host to run the software to predict the disease based on the data of the symptoms using the SVM algorithm, to record patient data, to record prediction results and plan of treatments data and also to centrally manage the machine algorithm module so that at the time of revision or updating the system, the mobile application does not require changes in many aspects. The result is then sent to a doctor to be validated. Furthermore, the doctor should give a recommendation or a treatment plan for the disease to the CHC. The CHC sends the result and the treatment plan to the patient. Based on that recommendation, the patient can conduct self-treatment using recommended medicines or come to the CHC for getting medicines if they do not have it. Fig. 4 shows the sequence diagram of the application which provides a preview of all events in the use case diagram. The advantage obtained by the enrolled patients is that they could get a prompt result regarding their disease prediction. They could also obtain the treatment from CHC by only inputting the symptoms of the disease.

The system adopts the client/server (CS) structure which uses Apache as the web server, PHP as server side scripting language, MySQL (MariaDB) as database and Bootstrap as front-end of the application as shown in Fig. 5. On the server-side, XAMPP for Windows ver 7.3.4 was installed on a PC with Intel i5 Microprocessor and 8 GB RAM to provide web server package containing Apache 2.4.39, MariaDB 10.1.38 and PHP 7.3.4. In the same PC, Php-ml [16] was installed using Composer, a package manager application to install PHP library dependencies. Php-ml is a machine learning library for PHP. In addition, Gammu 1.39.0 was deployed in the same PC for SMS Daemon. Meanwhile, on the client-side, the mobile application was developed using Android Studio and installed on a mobile phone with Android Nougat as the operating system. The user interface provides the data of the symptoms to the SMS manager. The SMS manager sent the data of the symptoms to SMS Daemon through the GSM network. When the data were received by Gammu, the data were then processed by PHP using

Fig. 3. Use case diagram of the intelligent mobile health monitoring system.

Table 1
Distribution of sex and age of the data.

Disease	Age	Male	Female
Typhoid	8-13	3	2
	18-25	0	2
	50-63	1	2
Cough	3	0	1
	18-32	0	4
	39-70	3	2
Diarrhoea	4–10	4	3
	19	2	0
	64	0	1
	Total	13	17

Php-ml. Php-ml used the SVM algorithm to determine to which group the data belongs. The classification result was recorded on the host PC to be accessed and validated by the doctor. The response of the classification result from doctor and CHC is recorded to MySQL database by PHP and sent back to the patient using Gammu. The mobile application Graphical User Interface (GUI) to input patient data and symptoms are shown in Fig. 6a and b. The data are sent to the server as the following format; the patient data are sent as it is, while the symptoms are sent using number 1 and 0. If the patient chooses "yes" it is marked with 1 and if the patient chooses "no" it is marked with 0. Each item is separated by comma. For instance, the formatted result was 98776534,Arman,56,M,65,165,2,1,1,0,1,0,1,0,1,1. The result and plan of treatment were sent back to the patient as an SMS as shown in Fig. 6c.

Table 1 shows the data distribution of the patients. It shows that diarrhoea patients were dominated by children and cough patients were dominated by adults. While typhoid patients were a combination of children and adults. Fig. 7 shows the bar graph of the symptoms or features gathered from the patients. The graph shows the distribution of the symptoms or features for each disease. There were 3 classes (Typhoid, Cough, Diarrhoea) and each class has 10 patients. For instance in cough (SYMCOUGH) attribute, there were 20 patients had a cough and 10 patients who did not have a cough; for headache (SYMHA) attributes, 9 patients had a headache and 21 patients did not have a headache.

Php-ml used C-SVC type for classification using SVM. WEKA data mining software [17] and LibSVM [18] were used to obtain the best parameters of C-SVC type. Three types of the kernel were tested; linear, polynomial and RBF. In addition, two degrees and gamma (γ) were applied on polynomial and RBF kernel. Table 2 shows the result of the classification. It indicates that in this particular application, the RBF kernel provided better performance as can be seen that correct classified data were better compared to linear and polynomial kernels. Compared to the polynomial kernel using various degrees and gammas (γ), RBF kernel also provided better correct and incorrect classified data which were 80% and 20%. The table also shows that Relative Absolute Error (RAE) of RBF Kernel provided smaller values compared to the linear and polynomial kernel which was 60.74%. Receiver Operating Characteristic (ROC) curve revealed the performance of the classification model. Thus, area under the ROC curve also determines the performance of the classification model. The bigger the value of ROC area under curve, the better the classification model. As shown in Table 2, the area under ROC curve of RBF kernel provided better result compared to the linear and polynomial kernel. Therefore, the RBF kernel with γ 0.0 and degree 2 was chosen as the kernel.

4. Discussion

The SMS service was used to send the data of the symptoms to the server to ensure that communication can be well established and without relying on the Internet connection. The problem arose when it comes to formatting the text as a report; it needs special attention from the user to format it. The mobile phone application was designed to use a GUI based on yes/no questions to convert the user input as a formatted text and to send it to a provider for classification processing. The mobile phone application provides a simple way to input the data of the symptoms and to avoid mistakes of formatting the text. The data will be analysed using the SVM algorithm to predict the disease to help doctors get the disease prediction sooner. Furthermore, the symptoms and the prediction result were accessed by the doctor to be validated. In case the system fails to predict, the data are analysed by the doctor and the prediction result and plan of treatment were also submitted to the server to be forwarded to the patient by CHC using SMS. These results and plans of treatment will be stored in a database on the server as patient's medical records for the benefit of seeing the history of the disease that has been experienced by the patient. In addition, this medical record can also be used to monitor the spread of disease and can also be used by decision-makers at the central level regarding health services in an area.

In this study, the SVM was able to classify three diseases. Nine symptoms gathered from typhoid, cough and diarrhoea patients were almost similar. Some symptoms such as cough and fever belong to all of these diseases. Based on these symptoms, the SVM created a machine learning model. This model is used to predict diseases. Therefore, the selection of the kernels and parameters greatly determines the prediction performance [19]. In some cases, using different parameters provide a similar result. One of these best parameters has been chosen which is RBF, degree 2 and γ 0.0. However, a relatively large amount of sample patient data is needed to find a better pattern in the training data which results in the machine learning model accuracy. A sufficient amount of patient data can improve predictive performance.

Before the application is used, a pilot project must be conducted. The pilot project should be conducted in a remote or rural area to assess and evaluate the system by considering several aspects such as security, integration to national healthcare programs, metric the real benefit of the mobile phone application and sustainability [20].

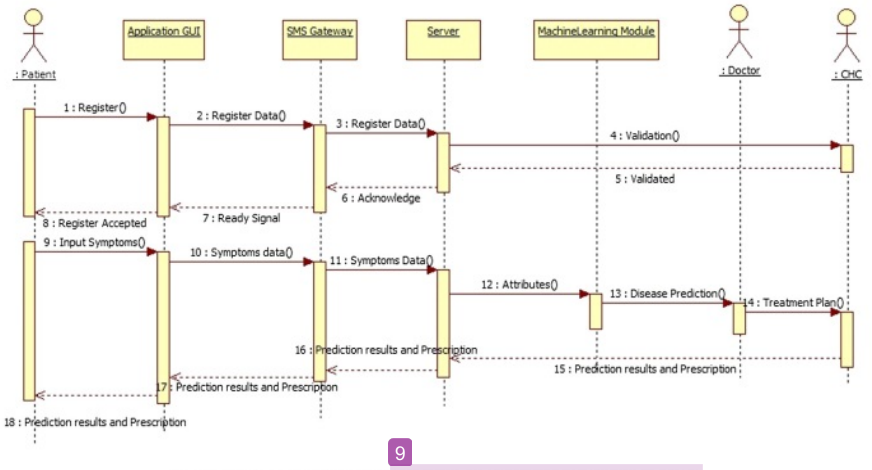


Fig. 4. Sequence diagram of the intelligent mobile health monitoring system.

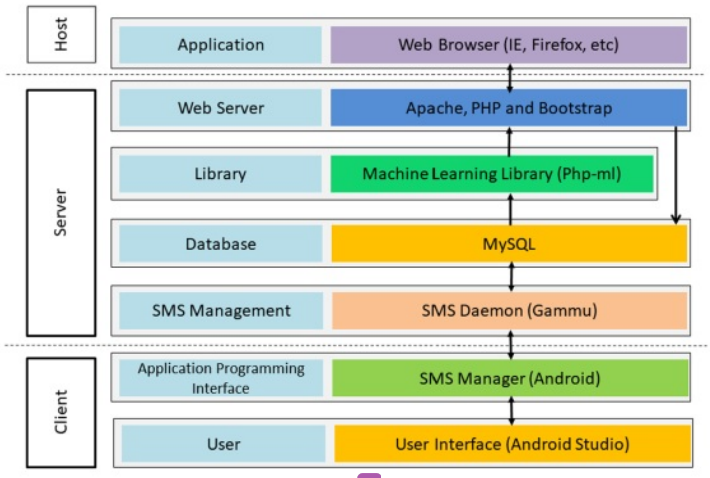
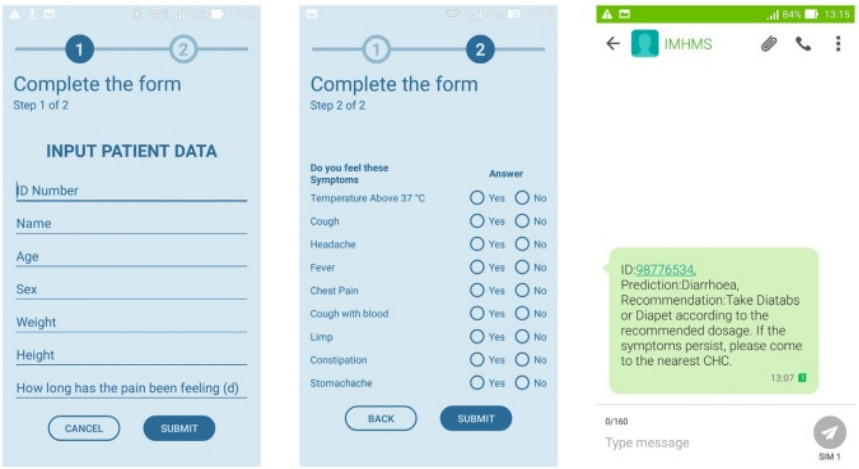


Fig. 5. Software architecture of the proposed intelligent mobile health monitoring system.



(a) (b) (c)

Fig. 6. The mobile application GUI forms for (a) input patient data (b) input symptoms (c) prediction results.

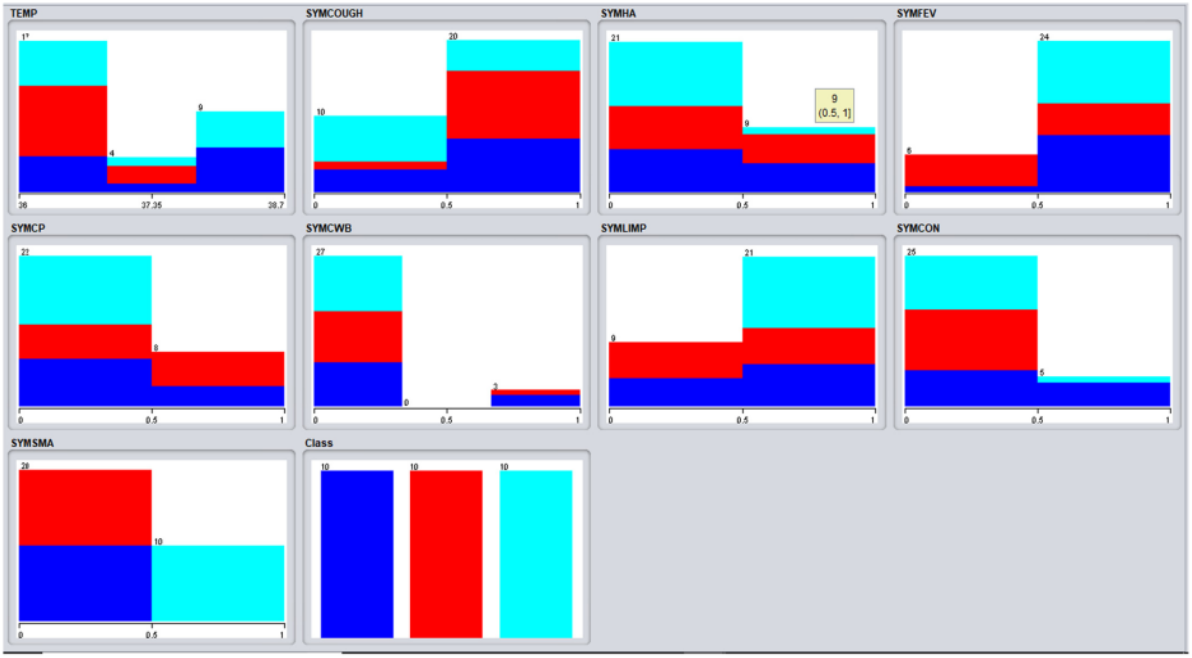


Fig. 7. Raw data figures of 30 patients with their symptoms based on their disease.

Table 2
Results of 3 types of kernel and different parameters to find the best classification result.

Classified	Linear	Polynomial				Radial Basis Function			
		$D=3$		$D=2$		$D=3$		$D=2$	
		$\gamma=.0$	$\gamma=.1$	$\gamma=.0$	$\gamma=.1$	$\gamma=.0$	$\gamma=.1$	$\gamma=.0$	$\gamma=.1$
Correctly (%)	76.7	76.7	76.7	73.3	76.7	80	76.7	80	80
Incorrectly (%)	23.3	23.3	23.3	26.7	23.3	20	23.3	20	20
RAE(%)	61.25	60.77	60.76	62.27	62.35	60.74	61.92	60.74	60.74
ROC Area	0.837	0.830	0.830	0.830	0.827	0.903	0.895	0.903	0.903

5. Conclusions

A mobile health monitoring system for telehealth has been designed and created. The mobile phone application was designed to ease people in rural and remote areas to obtain health care. Typhus, cough, and diarrhea were tested because these are some of the common diseases that many people suffer. Nine symptoms of diseases from 30 patients were gathered and marked for classification purposes. The developed system was easy to use since it uses a simple GUI to minimize errors when data entry is carried out. In addition, this system also does not require expensive connection fees because it uses SMS where the costs are inexpensive. The use of machine learning helps doctors predict the diseases suffered by patients so that they are expected to provide fast and precise services to patients in need. Overall, the system works properly. However, this application still has to be evaluated in the field to see the drawbacks and to get input from several related parties.

Human and animal rights

Not applicable.

Informed consent and patient details

The authors declare that this report does not contain any personal information that could lead to the identification of the patient(s).

The authors declare that they obtained a written informed consent from the patients and/or volunteers included in the article. The authors also confirm that the personal details of the patients and/or volunteers have been removed.

Funding

This work has been supported by: Ministry of Research, Technology and Higher Education of Republic Indonesia No. 277/UN36.9/PL/2017.

Author contributions

All authors attest that they meet the current International Committee of Medical Journal Editors (ICMJE) criteria for Authorship.

CRediT authorship contribution statement

Y.A. Djawad: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Soft-

ware, Visualization, Writing - original draft, Writing - review & editing. **S. Suhaeb:** Conceptualization, Investigation, Resources, Validation, Writing - original draft. **Ridwansyah:** Conceptualization, Data curation, Formal analysis, Investigation, Software, Validation, Writing - original draft. **H. Jaya:** Conceptualization, Methodology, Resources, Visualization, Writing - original draft. **Fathahillah:** Conceptualization, Data curation, Software, Validation. **Saharuddin:** Conceptualization, Formal analysis, Funding acquisition, Project administration, Supervision, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial or personal relationships that could be viewed as influencing the work reported in this paper.

Acknowledgement

This research is supported by the Indonesian Research, Technology and Higher Education Ministry through Decentralisation Scheme with contract No: 277/UN36.9/PL/2017.

References

[1] Kartika LW. Facilitators and barriers to health workforce retention in rural and remote setting of Indonesia: a literature review. KnE Life Sci 2018:140–57. https://doi.org/10.18502/kls.v4i10.3716, https://knepublishing.com/index.php/KnE-Life/article/view/3716.

[2] Efendi F. Health worker recruitment and deployment in remote areas of Indonesia. Rural Remote Health 2012;12:2008.

[3] Firdaus A, Efendi F, Hadisuyatmana S, Aurizki GE, Khatijah L. Factors influencing the intention of Indonesian nursing students to work in rural areas focus on Asia Pacific. Fam Med Commun Health 2019;7:144. https://doi.org/10.1136/fmch-2019-000144.

[4] Puspitasari IM, Soegijoko S. Development of an integrated e-prescription system with adverse drug events alert for community health center in Indonesia. In: Med-e-Tel international conference of e-health and telemedicine; 2009.

[5] Soegijoko S, Aridarma A, Puspitasari IM. Preliminary evaluation on experimental m-health system to support mother and child care promotion for community health centers in Indonesia. In: SHOPS and mHealth alliance hold online conference: using mobile technologies to improve family planning, maternal health and newborn services in the developing world; 2010.

[6] Hariyati TS, Sahar J. Perceptions of nursing care for cardiovascular cases, knowledge on the telehealth and telecardiology in Indonesia. Int J Collab Res Int Med Publ Health 2012;4(2):116–28.

[7] Sandi G, Nugraha IGBB, Supangkat SH. Mobile health monitoring and consultation to support hypertension treatment. In: International conference on ICT for smart society; 2013. p. 1–5.

[8] Jatmiko W, Ma'sum MA, Isa SM, Imah EM, Rahmatullah R, Wiweko B. Developing smart telehealth system in Indonesia: progress and challenge. In: 2015 international conference on advanced computer science and information systems; 2015. p. 29–36.

[9] Sutjiredjeki E, Soegijoko S, Mengko TLR, Tjondronegoro S, Astami K, Muhammad Suherman HU. Application of a mobile telemedicine system with multi communication links for disaster reliefs in Indonesia. In: World congress on medical physics and biomedical engineering. Berlin, Heidelberg: Springer; 2009. p. 344–7.

[10] Rahmatullah R, Habibie I, Mursanto P, Isa SM. TERRAIN: fetal growth telehealth system based on 2D fetal head image using randomized hough transform. J Ilmu Komput Inf 2014;7(1):37–43. https://doi.org/10.21609/jiki.v7i1.255. https://scholar.ui.ac.id/en/publications/terrain-fetal-growth-telehealth-system-based-on-2d-fetal-head-ima.

[11] Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995;20(3):273–97. https://doi.org/10.1007/BF00994018.

[12] Karamizadeh S, Abdullah SM, Halimi M, Shayan J, Rajabi Mj. Advantage and drawback of support vector machine functionality. In: 2014 international conference on computer, communications, and control technology (I4CT); 2014. p. 63–5.

[13] Statnikov A, Aliferis CF. Are random forests better than support vector machines for microarray-based cancer classification? AMIA Annual Symp Proc 2007:686–90.

[14] Statnikov A, Wang L, Aliferis CF. A comprehensive comparison of random forests and support vector machines for microarray-based cancer classification. BMC Bioinform 2008;9:319. https://doi.org/10.1186/1471-2105-9-319.

[15] Kavakiotis I, Tsave O, Salifoglou A, Maglaveras N, Vlahavas I, Chouvarda I. Machine learning and data mining methods in diabetes research. Comput Struct Biotechnol J 2017;15(Suppl C):104–16. https://doi.org/10.1016/j.csbj.2016.12.005.

[16] Kondas A. Php-ml - machine learning library for php. https://github.com/php-ai/php-ml, 2018.

[17] Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA data mining software: an update. SIGKDD Explor 2009;11(1):10–8.

[18] Chang C-C, Lin C-J. LIBSVM: a library for support vector machines. ACM Trans Intell Syst Technol 2011;2:27. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[19] Huang S, Cai N, Pacheco PP, Narandes S, Wang Y, Xu W. Applications of support vector machine (SVM) learning in cancer genomics. Cancer Genomics Proteomics 2017;15(1):41–51. https://doi.org/10.21873/cgp.20063. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5822181/.

[20] Gutierrez MA, Moreno RA, Rebelo MS. Chapter 3 - Information and communication technologies and global health challenges. In: de Fátima Marin H, Massad E, Gutierrez MA, Rodrigues RJ, Sigulem D, editors. Global health informatics. Academic Press; 2017. p. 50–93.

paper 3

ORIGINALITY REPORT

12%	9%	9%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	S. Chakraborty, C. Das, K. Ghoshal, M. Bhattacharyya, A. Karmakar, S. Chattopadhyay. "Low Frequency Impedimetric Cell Counting: Analytical Modeling and Measurements", IRBM, 2020 Publication	3%
2	Submitted to An Giang University Student Paper	1%
3	insist.unila.ac.id Internet Source	1%
4	D. Bibbo, S. Gabriele, A. Scorza, M. Schmid, S.A. Sciuto, S. Conforto. "A Novel Technique to Design and Optimize Performances of Custom Load Cells for Sport Gesture Analysis", IRBM, 2019 Publication	1%
5	dspace.cvut.cz Internet Source	1%
6	www.mdpi.com Internet Source	1%

7	"Recent Advances on Soft Computing and Data Mining", Springer Science and Business Media LLC, 2020 Publication	<1%
8	coek.info Internet Source	<1%
9	"Health Information Science", Springer Science and Business Media LLC, 2013 Publication	<1%
10	Submitted to Roehampton University Student Paper	<1%
11	Submitted to Universiti Teknikal Malaysia Melaka Student Paper	<1%
12	repository.tudelft.nl Internet Source	<1%
13	www.apachefriends.org Internet Source	<1%
14	Guangyou Xu. "Automatic image annotation by using concept-sensitive salient objects for image content representation", Proceedings of the 27th annual international conference on Research and development in information retrieval - SIGIR 04 SIGIR 04, 2004 Publication	<1%

15	doi.org Internet Source	<1%
16	"Data Science and Analytics", Springer Science and Business Media LLC, 2018 Publication	<1%
17	www.sevgigurbuz.com Internet Source	<1%
18	repository.ubaya.ac.id Internet Source	<1%
19	epdf.tips Internet Source	<1%
20	norma.ncirl.ie Internet Source	<1%
21	Wiwin - Suwarningsih. "The concept of Cloud Computing Services For the application of e-health Education", International Journal of Cloud Computing and Services Science (IJ-CLOSER), 2013 Publication	<1%
22	Strain, P.. "Design and implementation of a program to monitor ocean health", Ocean and Coastal Management, 2002 Publication	<1%
23	A Dirpan, R Latief, A Syarifuddin, A N F Rahman, R P Putra, S H Hidayat. " The use of	<1%

colour indicator as a smart packaging system for evaluating mangoes Arummanis (L. var. Arummanisa) freshness ", IOP Conference Series: Earth and Environmental Science, 2018
Publication

---

24 "Advances and Trends in Artificial Intelligence. From Theory to Practice", Springer Science and Business Media LLC, 2019 <1%
Publication

---

25 C. Caredda, L. Mahieu-Williame, R. Sablong, M. Sdika, J. Guyotat, B. Montcel. "Real Time Intraoperative Functional Brain Mapping Based on RGB Imaging", IRBM, 2020 <1%
Publication

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	On		

paper 4

by Yasser Djawad

Submission date: 24-Feb-2021 12:14PM (UTC+0500)
Submission ID: 1516870403
File name: Yasser_Sensing_and_Biosensing_Research.pdf (1.3M)
Word count: 5086
Character count: 25322

Sensing and Bio-Sensing Research 23 (2019) 100269

Contents lists available at ScienceDirect

Sensing and Bio-Sensing Research

journal homepage: www.elsevier.com/locate/sbsr

The application of detrended fluctuation analysis to assess physical characteristics of the human cell line ECV304 following toxic challenges

Yasser Abd Djawad[a,*], David Attwood[b], Janice Kiely[b], Richard Luxton[b]

[a] Electronics Department, Universitas Negeri Makassar, Indonesia
[b] Institute of Biosensing Technology, University of the West of England, UK

ARTICLE INFO

Keywords:
Cell toxicity
Impedance spectroscopy
Morphological size changes

ABSTRACT

In this paper we present a non-contact, impedance-based sensor system capable of characterizing the toxic response of cells to three different types of toxin. ECV304 cells were treated with 1 mM Hydrogen peroxide, 5% Dimethyl Sulfoxide, and 10 μg/ml saponin. Impedance spectroscopy was performed over a 2 h period on the cells within a commercial cell growth chamber, positioned on a pair of measurement electrodes, at frequencies between 200 and 830 kHz at 10 kHz intervals. Analysis of the impedance data was undertaken using the feature-extraction technique, Detrended Fluctuation Analysis (DFA). DFA scales the autocorrelation of a non-stationary signal, such as those generated using impedance spectroscopy for cytotoxicity testing. The correlation between the average fluctuation of the signal, F(n) (and scaling exponent, α) and a measurement of the cell size from image analysis was evaluated. The results showed that F(n) and α were strongly related to the changes of the morphological size of the cells. The results demonstrated that non-contact impedance spectroscopy, coupled with DFA can be used to monitor cell size in real time.

1. Introduction

A number of biochemical tests, based on the measurement of colour, are widely used to evaluate the toxicity of a drug or chemical by assessing viability of cells challenged with a potential toxin. These traditional approaches involve a number of steps and consequently they are inappropriate for real time or continuous monitoring. A number of techniques based on electrical measurements have been developed by researchers to study and monitor cell behaviour and viability [1–4]. Impedance spectroscopy is one such method and, because it can be miniaturised and gives instantaneous data, it has the potential be applied for the continuous monitoring of cells in culture [5–7]. Cytotoxic changes in cells have been demonstrated using impedance spectroscopy as a result of exposure to toxins and also viruses [1,5].

Classical methods for analysing data from impedance spectroscopy to study behaviour of the biological systems have been reported by many researchers. Commonly these methods used to analyse the impedance signal generated by the cells focus on either time domain or frequency domain analysis [6,8–12]. However, the analysis of the time or frequency domains alone can lead to an oversimplified understanding of the cells response to toxins. Therefore, Time-Frequency Representations (TFRs) have been applied to give a more complete representation of the data generated from monitoring toxin related changes in cells [13]. Short Time Fourier Transform (STFT) was used to discriminate three types of cells (Jurkat, ECV304 and Caco2) [14]. This work demonstrated that each type of cell generated a unique signal and STFT was able to distinguish the characteristics of those signals. Furthermore, another TFR technique, the Discrete Wavelet Transform (DWT) has been used to extract features from Electroencephalogram (EEG) signals [13,15]. However, the challenge of these TFRs techniques is that a very large data set is generated; this then requires the application of reducing techniques in order to create a smaller, more manageable data package. Wavelet Package Decomposition was also used to characterise impedance spectroscopy signals from ECV304 cells treated with Hydrogen peroxide (H_2O_2) [16]. The signals from the cells were characterised using the coefficients derived from the decomposition algorithm, which related to the cell morphology. The advantage of this approach is that it gave a systematic process for characterisation of the large data sets obtained from cell monitoring experiments using impedance spectroscopy giving a route to automated data analysis.

Other data processing techniques also provide alternatives to analysing large data sets obtained from biomedical experiments. One such technique is Detrended Fluctuation Analysis (DFA), which is a statistical method for scaling long-range correlations, Advantages of DFA over other methods (e.g. spectral analysis and Hurst analysis) are that it permits the detection of intrinsic self-similarity embedded in a

* Corresponding author.
E-mail address: yasser.djawad@unm.ac.id (Y.A. Djawad).

https://doi.org/10.1016/j.sbsr.2019.100269
Received 10 November 2018; Received in revised form 21 February 2019; Accepted 27 February 2019

seemingly non-stationary time series, and avoids the spurious detection of apparent self-similarity, which may be an artefact of extrinsic trends. DFA has been used to analyse biosignals generated by cells and organs, for example, to investigate the fluctuation properties of an ion channel [17] and to model patterns of breathing [18]. In another example, DFA was used in biomedical signal processing for the discrimination of heart rate variability between children with type 1 diabetes with micro-albuminuria and healthy children [19].

In this study, a cytotoxicity test was developed using ECV304 cells to investigate the response to three different toxins and impedance data was analysed using DFA. Key parameters created using the DFA method, the fluctuation average F(n) and the scaling exponent, α, were correlated with cell size changes as observed via microscopy. The toxins and their concentrations were selected based on their physical properties and mechanism of toxicity. Toxins used were H_2O_2, a potent oxidising agent; Dimethyl Sulfoxide (DMSO), a potent cell differentiating agent; and saponin, a potent membrane permeabilizing agent. H_2O_2 is used in the food and pharmaceutical industries and is reported to cause cell damage through direct oxidation of lipids, proteins and DNA [20,21]. The main applications for DMSO are found in the food, pharmacy and agrochemicals sectors. A previous study showed that DMSO changed the morphology of cells and, depending upon the concentration, reduced the cell viability [22]. Saponins are mostly used in the pharmaceutical and cosmetic industry. Saponin is a detergent-like molecule that is able to permeabilise the plasma membrane of the cells. Saponin penetrates the lipid bilayer structure of the cells and binds with the cholesterol to lyse cells [23].

2. Materials and methods

2.1. Cells preparation

ECV304 cells, from the European Collection of Cell Cultures (Public Health England) were seeded at a density of 3×10^5 cells/mL in a 75 cm^3 flask in 12 mL M199 media (Gibco) supplemented with 10% foetal calf serum and 2 mL L-glutamine and maintained in a humidified incubator at 37 °C with 5% CO^2. The medium was changed every 3 days to feed the cells and monitored daily by microscope to check for confluence, at which point the cell monolayer covers 75–80% of flask surface. The 70–80% confluence was estimated by observing and comparing space occupied by the cells with the area that is not occupied by the cells. When confluence was achieved, the cell layer was washed with Phosphate Buffered Saline (PBS) followed by 1 mL of Trypsin-EDTA solution (0.05% porcine trypsin, 0.2 g/L EDTA) followed by incubation at 37 °C for approximately 5 min to allow the cells to detach from the flask surface. Once the cells had detached, the trypsin was neutralised by the addition of 2 mL of growth medium and the cell density calculated by the Trypan Blue Exclusion method. The cell suspension was diluted to a density of 3×10^5 cells/mL and 2 mL of cell suspension added to a chamber in a 2-chamber tissue culture plate (Nunc LAB-TEK II Chambered Coverglass, USA) and maintained at 37 °C and 5% CO^2 until cell monolayer covers 100% of flask surface. Cytotoxicity testing was performed by adding 1 mM H_2O_2, 5% DMSO and 10 µg/ml saponin. Images were recorded using Nikon phase contrast microscope (Nikon Instruments Inc., USA) at a magnification of 200 at 1, 80 and 120 min for each toxic challenge.

2.2. Instrumentation

Impedance measurements were conducted using the D patterned electrodes connected to a lock-in amplifier (LIA) [24]. The electrode was produced using PCB fabrication techniques and comprised a copper layer coated with 0.1 µm gold over 5 µm nickel. The diameter of each sensor was 22 mm, with 1 mm space between the counter and detecting electrodes, the total sensor area corresponding to the culture area of the cell chamber. The D patterned sensor was connected in series with a

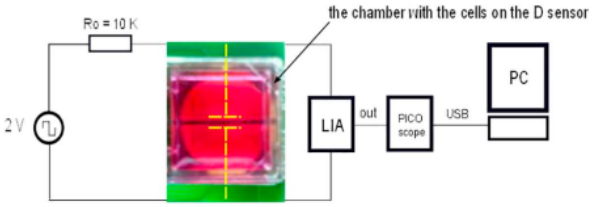


Fig. 1. Circuit diagram of the system and a photograph of the D sensors below the cell chamber.

10 KΩ resistor to limit the current and to achieve a large bandwidth for the RC circuit. Measurements were performed by connecting the lock-in amplifier in parallel with the sensors as shown in Fig. 1. The tissue culture chamber was placed on the D sensor such that the chamber containing the cells was accurately located above the sensor, but not in direct contact with the sensor. The base of the tissue culture chamber was made of borosilicate glass with a thickness of 0.15 mm and the culture area was 4.2 cm^2. It was critical that the base of the chamber incorporated thin glass to ensure that the electric field penetrated into the cell culture. During the experiments, the D sensors with the cell chamber were placed in the incubator. The input signal, was a 2 V square wave signal over a frequency range of 200 to 830 kHz and the measurement of voltage output from the LIA was facilitated using a PICOscope. A PICOScope is a real time PC-based digital oscilloscope, which has function of digital storage oscilloscope, spectrum analyser and signal generator (Pico Technology, UK). This frequency range was selected as it had been demonstrated that significant changes in the impedance spectra occurred on the application of a toxic challenge to ECV304 cells over this frequency range [25]. Data was acquired for each data set at 1, 80 and 120 min after either H_2O_2, DMSO or saponin had been added. The total data set from each experiment comprised of 64 blocks; sampled from 200 kHz to 830 kHz with an interval 10 kHz.

2.3. Detrended fluctuation analysis

DFA is a statistical technique for scaling long range correlations in a time series. It quantifies the complexity of signals by first least square fitting a line to an integrated time series and subtracting that fit (the mean). Subsequently the root mean square of the fluctuations around this fit is found and this is repeated for multiple time scales, as described below.

For a 1-D time series X(i), $i = 1, \ldots, N$, where N is the length of the time series. The integral of the time series can be computed using following formula:

$$y(k) = \sum_{i=1}^{k} (X(i) - X_{avg}) \tag{1}$$

where X_{avg} is the mean value of the time series $X(i)$. The integrated time series is divided into several windows with length n, the scale of the window size. In each window, a least square line is applied to obtain best approximation of data trend or linear approximation of the data. The average fluctuation $F(n)$ of the signal in the trend can be described as:

$$F(n) = \sqrt{\frac{1}{N}\left(\sum_{k=1}^{N} (y(k) - y_n(k))^2\right.} \tag{2}$$

Different values of n, the data scale, are selected and corresponding values of $F(n)$ are calculated. To analyse the result, a plot of log $F(n)$ versus log n is created and the slope of the data trend from this graph is determined as the scaling exponent, α. The scaling exponent represents the autocorrelation of the time series with following criteria:

1. $\alpha < 0.5$ anti-correlated signal.

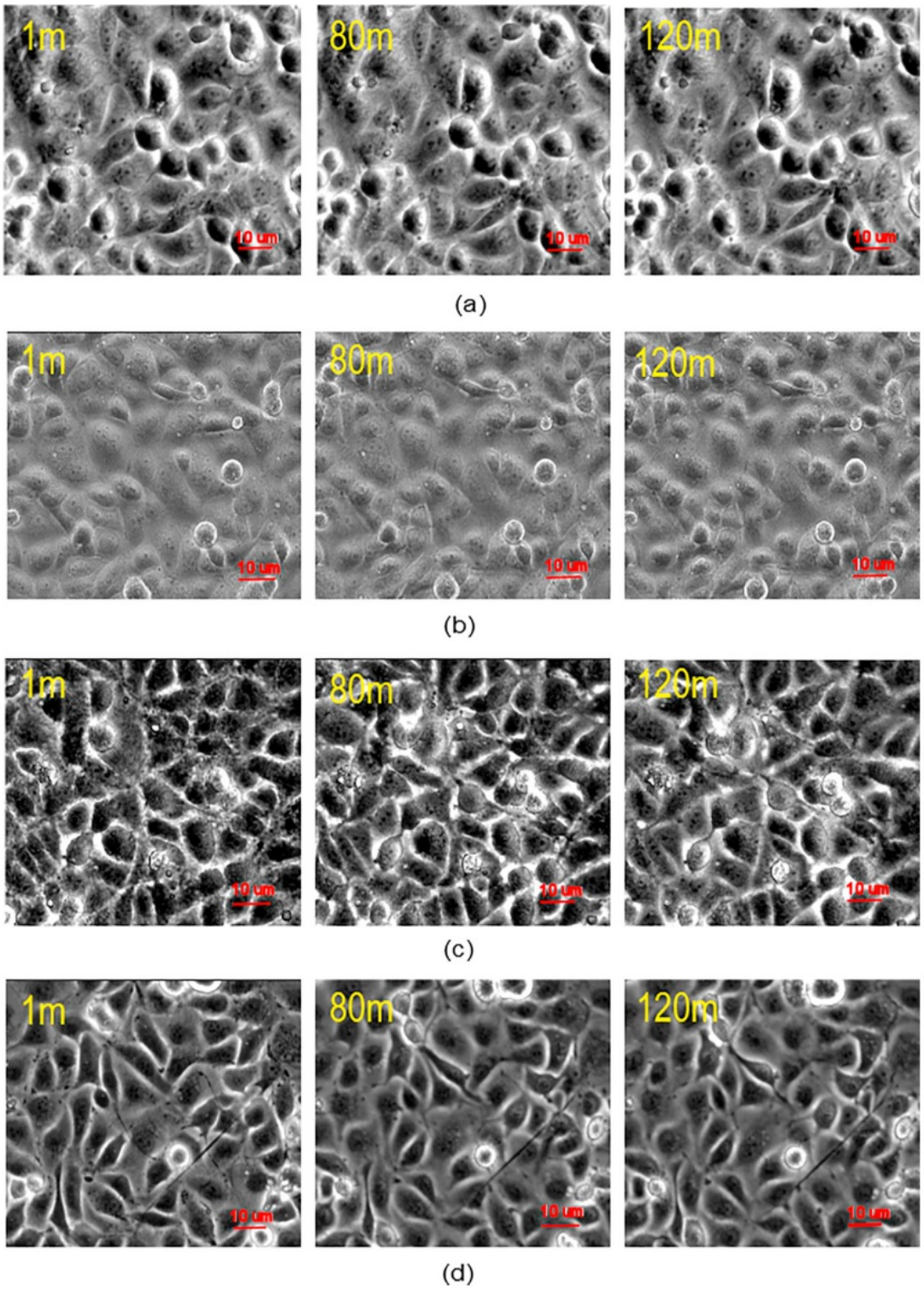


Fig. 2. Images of ECV304 cells with 3 type of toxins (a) Control cells (b) 1 mM H_2O_2 (c) 5% DMSO (d) 10 µg/ml saponin.

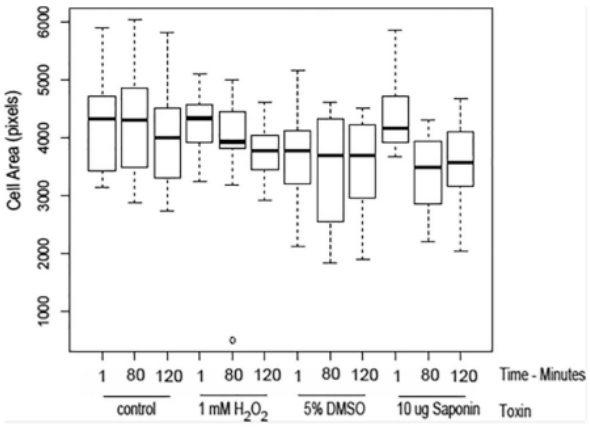


Fig. 3. Box and whiskers plot of cell size measurement using different types of toxin.

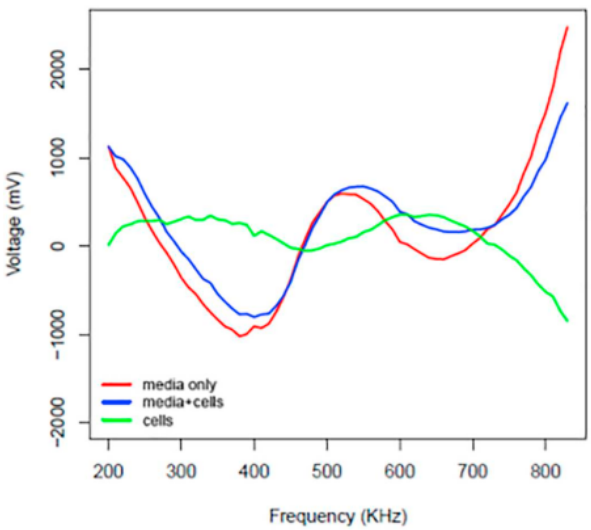


Fig. 4. IS spectra and IS difference spectra from cells and media.

2. $\alpha \simeq$ 0.5 uncorrelated signal (white noise).
3. $\alpha >$ positive autocorrelation in the time series.
4. $\alpha \simeq$ 1/f noise.
5. $\alpha \simeq$ 1.5 Brownian noise or random walk.

3. Results and analysis

3.1. Image analysis

Images were captured and cell sizes were measured using Image-J for 10 cells in each image at 1, 80 and 120 min for each type of toxin as shown in Fig. 2. The cell size was expressed as the measured area from the image expressed by pixel number.

Fig. 3 shows Box and Whiskers plots of the cell area derived from the images. For the control cells and those treated with DMSO there was no significant change in cell size over the 180 min the measurements were made. Cells treated with H_2O_2 showed a significant reduction in cell size across the three time points. For cells treated with saponin there was a significant reduction in cell size between 1 min and 80 min of exposure ($p < 0.01$) whereas there was not a significant difference in the cell size between 80 and 120 min exposure to saponin. This implies that by 80 min the cells had recovered from the stress associated with the addition of saponin.

3.2. Detrended fluctuation analysis

DFA was conducted on the impedance spectroscopy data obtained for control cells and those treated with H_2O_2, DMSO and saponin. The data used was the difference data obtained from subtracting the spectra of the media from that of the spectra of the media and cells, as shown in Fig. 4. The difference spectra were then segmented into 64 data batches sampled from 200 kHz to 830 kHz with an interval of 10 kHz.

Fig. 5 illustrates the average fluctuation F(n) for ECV304 cells following exposure to the three toxins, and control, for 1, 80 and 120 min for each scale, i.e. the number of segments the total data set is divided, each containing a number of 10 kHz data sets. In all cases the magnitude of the average fluctuation, F(n), increased with the scale. This is as expected because as the scale increases the number of data sets used in the analysis increases. For example, for the control sample, F(n) is ≈885 when a scale of 6 is used and F(n) ≈ 2795 for a scale of 24. For all the treatments at 1 min exposure there was a mean 1.85 fold increase in F(n) going from scale 6 to scale 12 and a mean 1.75 fold increase in F(n) going from scale 12 to 24. This was consistent across all the exposure times.

The range of F(n) did vary with the treatments. For the control cells the range of F(n) varied slightly across the exposure times was 5 for scale 6, 10 for scale 12 and 14 for scale 24. For cells treated with saponin, compared with the control treatment there was no significant change in the F(n) range at scales 6 and 12, but an increased range of 75 at scale 24. There was a significant difference in the F(n) range for cells exposed to H_2O_2 or DMSO compared with the control, at all times with the range reaching 200 for H_2O_2 treated cell and 280 for DMSO treated cells, for scale 24. This suggests that the data has higher levels of fluctuation in cells treated with toxins with DMSO.

The control cells showed that the values of F(n) on the scale 6 and the scale 12 tend to decrease at 80 min and increase at 120 min but at scale 24 there was a generalised decrease of F(n) values over all time periods as shown by Fig. 5a. When cells were exposed to H_2O_2, F(n) decreased with exposure time. This was seen at all scales as shown in Fig. 5b. Cells treated with 5% DMSO showed the greatest F(n) values and the greatest range of values. All the scales showed similar changes over time with a decrease in F(n) between 1 and 80 min followed by and increase between 80 and 120 min as seen in Fig. 5c. Fig. 5d shows the graph of cells +10 μg/ml saponin and depicts different patterns for each scale used. As noted above, the F(n) values were not significantly different from those obtained from the controls except at scale 24.

Increasing F(n) values indicate increasing fluctuation in the impedance spectra obtained from the cells. The toxins used are well known to cause cell disruption which can result in changes of cell size and alteration of the internal structure of the cells, both of which can affect the impedance of the cell. The results suggest that for cells treated with H_2O_2 or saponin there is a significant decrease in cell size which results in a reduction in the fluctuation of in the impedance signal. Treatment with DMSO resulted in the highest values of F(n) observed indicating the greatest level of fluctuation in the impedance signal derived from the cells. The F(24) values did not show a linear trend as did the other toxins which may reflect the lack of change in cell size over time. The higher values of F(n) may be a result of the greater variability in the cell size as indicated by the interquartile distance in Fig. 3. Although there was not a significant change in cell size the results suggest that 5% DMSO significantly disrupts cells by affecting the internal structure of the cells whilst not changing of cell size over time.

At scale 24 there is a correlation between the F(n) and cell size for cells treated with hydrogen peroxide and saponin, as shown in Fig. 3. In both cases the change in cell size over time is reflected in the change of F(n) over time. At scale 24 the control cells demonstrated a drop in F(n) over time but the magnitude of the drop was only 14. Although there

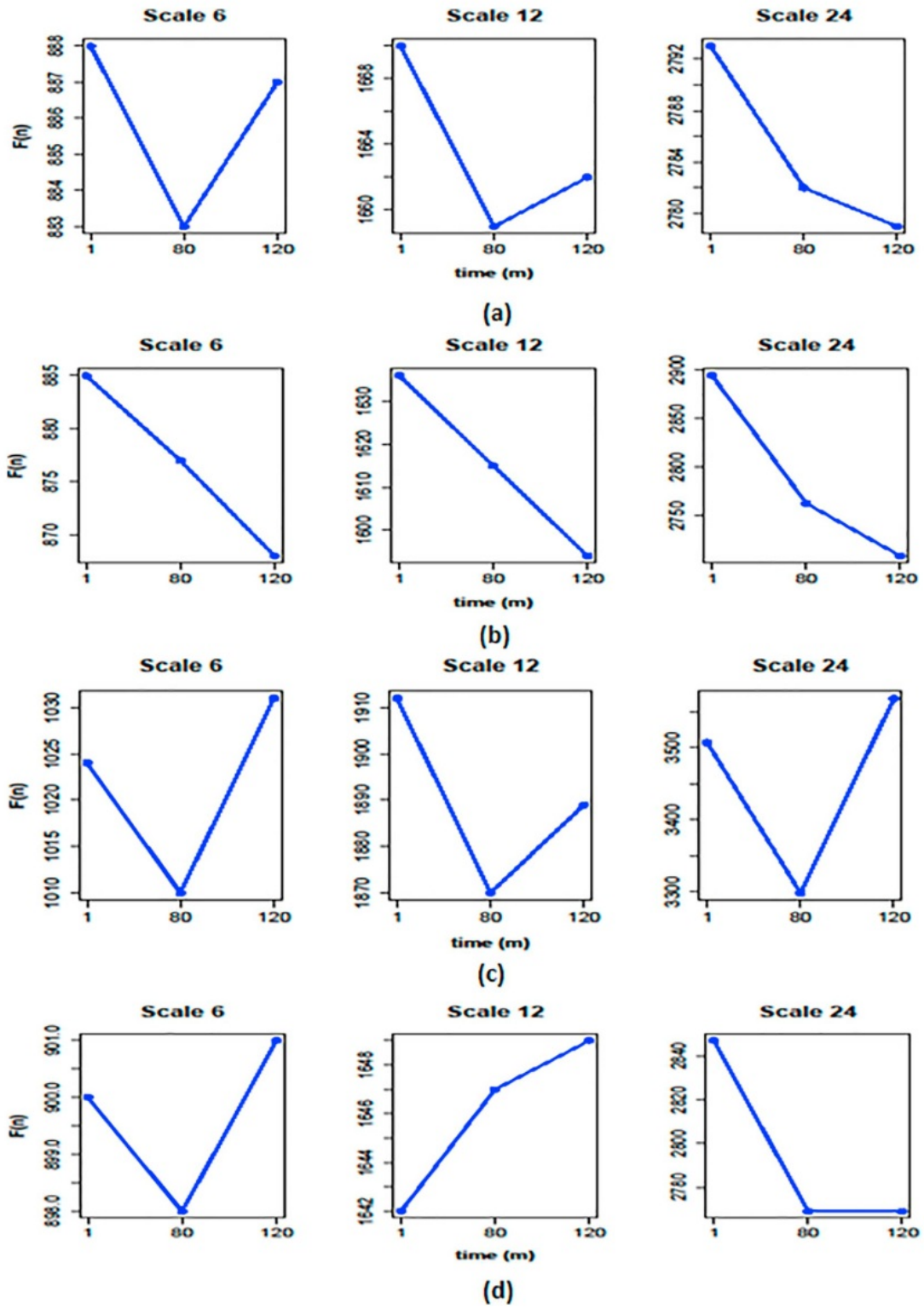


Fig. 5. The average fluctuation F(n) of (a) Control cells (b) cells + H_2O_2 (c) cells + DMSO (d) cells + saponin.

Table 1
Correlation coefficients between cell size and the average fluctuation of the signals and F(n) value calculated using DFA.

Treatments	F(n)		
	6	12	24
Control	0.1890	0.7034	0.9496
1 mM H_2O_2	0.9508	0.9607	0.9989
5% DMSO	0.6546	0.9989	0.7374
10 μg/ml saponin	0.4193	0.8660	0.9707

Table 2
Scaling exponent (α) of the four treatments of the ECV304 cells.

Treatments	1 m	80 m	120 m
Control	0.8326	0.8185	0.8181
1 mM H_2O_2	0.8597	0.8362	0.8296
5% DMSO	0.8897	0.8596	0.8914
10 μg/ml saponin	0.8361	0.8226	0.8203

was not a statistical difference in cell size over time there is a strong correlation between F(n) at scale 24 and the median cell size. Cells exposed to 5% DMSO demonstrated a drop in F(n) from 1 min to 80 min and then an increase in F(n) from 80 min to 120 min which is not reflected in the change in cell size. It is noted that the interquartile range increased significantly at 80 min compared with the range at 1 min. The interquartile range then decreased at 120 min. Correlation coefficients between cells size measured from microscopy images and F(n) are shown in Table 1. These clearly show the strong correlation between the mean cell size and the F(n) at scale 24. Cells exposed to DMSO show the greatest correlation at scale 12. These results imply that F(n) correlates well with cell size and reflects changes in cell size for certain toxins, in particular a highly significant correlation was observed for the effect of H_2O_2 on cell size over time.

In addition to measuring F(n), the scaling exponent (α) was calculated for the control and three treatments of ECV304 cells. The values of α at 1, 80 and 120 min are presented in Table 2. Firstly, all values were greater than 0.8, which in accordance with DFA theory suggests that the noise associated with the signal was not increasing as a function of the scale used. The pattern of change of values over time were markedly similar to the patterns of change seen with F(n) at the scale of 24 over the three time periods. This may indicate that a scale of 24 provides a more accurate result in comparison with the scales 6 or 12 and it is noted that the impedance data at the different time points showed positive autocorrelation, as all the scaling exponents, α, of the time series were above 0.8.

A strong correlation between cell size and the scaling exponent, α, was found for all the cell treatments: Control = 0.8778, H_2O_2 = 0.994, DMSO = 0.8412, saponin = 0.9295. This is as expected as there is the strong similarity between the change in α and F(24). Apart from the treatment with DMSO the correlation coefficients of time versus scaling exponent were lower than the corresponding correlation coefficients of time versus F(24).

4. Conclusions

In this study, ECV304 cells were challenged with three toxins to induce cytotoxic changes in the cells. The toxins used have different mechanisms of toxicity resulting in cell changes which result in changes in the cell size with or without changes to the internal organelles of the cells. Images of the cells were taken over 120 min and changes in size noted. Impedance spectra of the cells were collected at 1 min, 80 min and 120 min of exposure to the toxins. The application of DFA on the impedance spectra was investigated as a method that could rapidly detect cell toxicity.

The results of DFA were compared with the measurement of cell size resulting from the toxic-challenge. A strong positive correlation was observed with F(24). In addition, the scaling exponents (α) also provided strong positive correlation with the cell size measurements. Cells treated with 1 mM H_2O_2 demonstrated the strongest correlation between changes measured by microscopy and changes measured by impedance data. Cells treated with 5% DMSO did not show a significant change in size and F(n) did not show a linear trend with time at all scales. The very high values compared with the control cell suggest that the cytotoxic effects of DMSO involve changes to the cell organelles.

The work reported suggests that an automated measure of the health of cells could be developed using impedance spectroscopy coupled with signal analysis using the DFA technique. Monitoring of the magnitude and the change in F(n) over time could indicate different cytotoxic effects on the cells in terms of cell size or changes in cell organelles. More work using a wider range of toxins could lead to the development of an algorithm that could accurately determine the effect of a toxin on the cell and aid the classification of cytotoxic effects of different chemicals.

Conflict of interest

None.

References

[1] C. Caviglia, K. Zr, S. Canepa, M. Carminati, L.B. Larsen, R. Raiteri, T.L. Andresen, A. Heiskanen, J. Emnus, Interdependence of initial cell density, drug concentration and exposure time revealed by real-time impedance spectroscopic cytotoxicity assay, Analyst 140 (10) (2015) 3623–3629, https://doi.org/10.1039/C5AN00097A URL https://pubs.rsc.org/en/content/articlelanding/2015/an/c5an00097a.

[2] R.Y.A. Hassan, M.M. Mekawy, P. Ramnani, A. Mulchandani, Monitoring of microbial cell viability using nanostructured electrodes modified with graphene/alumina nanocomposite, Biosens. Bioelectron. 91 (2017) 857–862, https://doi.org/10.1016/j.bios.2017.01.060 URL http://www.sciencedirect.com/science/article/pii/S0956566317300593.

[3] O. Salyk, J. Vteek, L. Omasta, E. afakov, S. Sttesk, M. Vala, M. Weiter, Organic electrochemical transistor microplate for real-time cell culture monitoring, Appl. Sci. 7 (10) (2017) 998, https://doi.org/10.3390/app7100998 URL http://www.mdpi.com/2076-3417/7/10/998.

[4] N. Gaio, A. Waafi, M.L.H. Vlaming, E. Boschman, P. Dijkstra, P. Nacken, S.R. Braam, C. Boucsein, P.M. Sarro, R. Dekker, A multiwell plate organ-on-chip (OOC) device for in-vitro cell culture stimulation and monitoring, IEEE Micro Electro Mech. Syst. (MEMS) 2018 (2018) 314–317, https://doi.org/10.1109/MEMSYS.2018.8346549.

[5] M.S. Cheng, S.H. Lau, K.P. Chan, C.-S. Toh, V.T. Chow, Impedimetric cell-based biosensor for real-time monitoring of cytopathic effects induced by dengue viruses, Biosens. Bioelectron. 70 (2015) 74–80, https://doi.org/10.1016/j.bios.2015.03.018.

[6] E. Sarr, M. Lecina, A. Fontova, F. Gdia, R. Brags, J.J. Cair, Real-time and on-line monitoring of morphological cell parameters using electrical impedance spectroscopy measurements, J. Chem. Technol. Biotechnol. 91 (6) (2016) 1755–1762, https://doi.org/10.1002/jctb.4765 URL https://onlinelibrary.wiley.com/doi/abs/10.1002/jctb.4765.

[7] D. Seidel, J. Obendorf, B. Englich, H.-G. Jahnke, V. Semkova, S. Haupt, M. Girard, M. Peschanski, O. Brstle, A.A. Robitzki, Impedimetric real-time monitoring of neural pluripotent stem cell differentiation process on microelectrode arrays, Biosens. Bioelectron. 86 (2016) 277–286, https://doi.org/10.1016/j.bios.2016.06.056.

[8] C. Dalmay, M. Cheray, A. Pothier, F. Lallou, M.O. Jauberteau, P. Blondy, Ultra sensitive biosensor based on impedance spectroscopy at microwave frequencies for cell scale analysis, Sensors Actuators A Phys. 162 (2) (2010) 189–197, https://doi.org/10.1016/j.sna.2010.04.023 URL http://www.sciencedirect.com/science/article/pii/S0924424710001998.

[9] L. Wang, L. Wang, H. Yin, W. Xing, Z. Yu, M. Guo, J. Cheng, Real-time, label-free monitoring of the cell cycle with a cellular impedance sensing chip, Biosens. Bioelectron. 25 (5) (2010) 990–995, https://doi.org/10.1016/j.bios.2009.09.012.

[10] M.A. Ahmad, Z.A. Natour, S. Attoub, A.H. Hassan, Monitoring of the budding yeast cell cycle using electrical parameters, IEEE Access 6 (2018) 19231–19237, https://doi.org/10.1109/ACCESS.2018.2820080.

[11] C. Slouka, D. J. Wurm, G. Brunauer, A. Welzl-Wachter, O. Spadiut, J. Fleig, C. Herwig, A novel application for low frequency electrochemical impedance spectroscopy as an online process monitoring tool for viable cell concentrations, Sensors (Basel, Switzerland) 16 (11). doi:https://doi.org/10.3390/s16111900. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5134559/

[12] J.B.J.H. van Duuren, M. Musken, B. Karge, J. Tomasch, C. Wittmann, S. Hussler, M. Brnstrup, Use of single-frequency impedance spectroscopy to characterize the growth dynamics of biofilm formation in pseudomonas aeruginosa, Sci. Rep. 7 (1) (2017) 5223, https://doi.org/10.1038/s41598-017-05273-5 URL https://www.

nature.com/articles/s41598-017-05273-5.
[13] S.E. Khaleelulla, P.R. Kumar, EEG signal analysis for mental states and conditions of human brain, Proceedings of 2nd International Conference on Micro-Electronics, Electromagnetics and Telecommunications, Lecture Notes in Electrical Engineering, Springer, Singapore, 2018, pp. 141–151, , https://doi.org/10.1007/978-981-10-4280-5_15.
[14] Z. Wang, J. Kiely, M. Nibouche, R.W. Luxton, Z. Wang, J. Kiely, M. Nibouche, R.W. Luxton, Impedimetric discrimination of cell types for use in a whole cell biosensor, Proceedings of the 10th World Congress on Biosensors, Elsevier, Shanghai, China, 2008URL http://www.sparksdesigns.co.uk/workingfiles/bluezulu/biosensors08/index.htm.
[15] V. Geethu, S. Santhoshkumar, An efficient FPGA realization of seizure detection from EEG signal using wavelet transform and statistical features, IETE J. Res. (2018) 1–11, https://doi.org/10.1080/03772063.2018.1491806.
[16] Y.A. Djawad, J. Kiely, M. Nibouche, P. Wraith, R. Luxton, Robust feature extraction from impedimetric signals using wavelet packet decomposition with application to cytotoxicity testing, IET Sci. Meas. Technol 6 (6) (2012) 456–463, https://doi.org/10.1049/iet-smt.2012.0009.
[17] T.-H. Lan, Z.-Y. Gao, A.N. Abdalla, B. Cheng, S. Wang, Detrended fluctuation analysis as a statistical method to study ion single channel signal, Cell Biol. Int. 32 (2) (2008) 247–252, https://doi.org/10.1016/j.cellbi.2007.09.001 URL http://www.sciencedirect.com/science/article/pii/S1065699507002338.
[18] B.F. BuSha, G. Banis, A stochastic and integrative model of breathing, Respir. Physiol. Neurobiol. 237 (2017) 51–56, https://doi.org/10.1016/j.resp.2016.12.012.
[19] A. Golińska, Detrended fluctuation analysis (DFA) in biomedical signal processing: selected examples, studies in logic, Grammar Rhetoric 29 (2012) 107–115.
[20] N. Nakayama, S. Yamaguchi, Y. Sasaki, T. Chikuma, Hydrogen peroxide-induced oxidative stress activates proteasomal trypsin-like activity in human U373 glioma cells, J. Mol. Neurosci. 58 (2) (2016) 297–305, https://doi.org/10.1007/s12031-015-0680-9.
[21] L. Zhou, Y. Peng, Q. Wang, Q. Lin, An ESIPT-based two-photon fluorescent probe detection of hydrogen peroxide in live cells and tissues, J. Photochem. Photobiol. B 167 (2017) 264–268, https://doi.org/10.1016/j.jphotobiol.2017.01.011.
[22] R. Pal, M.K. Mamidi, A.K. Das, R. Bhonde, Diverse effects of dimethyl sulfoxide (DMSO) on the differentiation potential of human embryonic stem cells, Arch. Toxicol. 86 (4) (2012) 651–661, https://doi.org/10.1007/s00204-011-0782-2.
[23] R. Malabed, S. Hanashima, M. Murata, K. Sakurai, Sterol-recognition ability and membrane-disrupting activity of Ornithogalum saponin OSW-1 and usual 3-O-glycosyl saponins, Biochim. Biophys. Acta 1859 (12) (2017) 2516–2525, https://doi.org/10.1016/j.bbamem.2017.09.019.
[24] S.K. Sengupta, J.M. Farnham, J.E. Whitten, A simple low-cost lock-in amplifier for the laboratory, J. Chem. Educ. 82 (9) (2005) 1399, https://doi.org/10.1021/ed082p1399.
[25] Y.A. Djawad, J. Kiely, P. Wraith, R. Luxton, Lock-in amplifier as a sensitive instrument for biomedical measurement: analysis and implementation, Indones. J. Electr. Eng. Comput. Sci 12 (10) (2014) 7214–7222, https://doi.org/10.11591/ijeecs.v12.i10.pp7214-7222.

paper 4

ORIGINALITY REPORT

10%	6%	7%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

MATCH ALL SOURCES (ONLY SELECTED SOURCE PRINTED)

1%

★ Jens Kastenhofer, Vignesh Rajamanickam, Julian Libiseller-Egger, Oliver Spadiut. "Monitoring and control of E. coli cell integrity", Journal of Biotechnology, 2021

Publication

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	On		

paper 4

GRADEMARK REPORT

FINAL GRADE

/0

GENERAL COMMENTS

Instructor

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

paper 5

by Yasser Djawad

Submission date: 24-Feb-2021 12:34PM (UTC+0500)
Submission ID: 1516878896
File name: ej_yasser_rev3.docx (272.48K)
Word count: 4735
Character count: 24334

DOI:10.4186/ej.year.vol.issue.pp

ENGINEERING JOURNAL

Article

2024Essential feature extraction of photoplethysmography signal of men and women in their 20s

1215Yasser Abd Djawad[1,a] , Andi Mu'nisa[1], Pangayoman Rusung[1], Abdi Kurniawan[1], Irma Suryani Idris[1] , and Mushawwir Taiyeb[1]

[1] Universitas Negeri Makassar, JL. AP Pettarani, Makassar 90222, Indonesia

E-mail: [a]yasser.djawad@unm.ac.id (Corresponding author)

Abstract.

This study aims to extract the essential features of Photoplethysmography (PPG) signal of men and women in healthy subjects using Power Spectral Density (PSD) and Detrended Fluctuation Analysis (DFA). A PPG instrument was used to obtain the PPG signal of 15 men and 15 women. Using PSD, four frequency bands were selected to divide the spectral component. The areas within the frequency bands relative to the total area were computed as features of the signals. Furthermore, using DFA, the average fluctuation F(w) was computed. The feature extraction using this technique produced 4 features from different windows. Hurst exponent was calculated to analyse the characteristics of the time series. For comparing the feature extraction techniques, Heart Rate (HR) and Peak to Peak Interval (PPI) were computed. Additionally, F and T tests for all techniques were computed to determine the differences between man and woman features that have been gathered using these two techniques. The results indicate that the features of PPG signals of men and women using PSD and DFA were significantly different. In order to evaluate the results, a clustering analysis was applied to the results using K-means clustering technique. The clustering plots show that the features were well distributed into the two groups.

Keywords: feature extraction, frequency bands, spectral component, average fluctuation

ENGINEERING JOURNAL Volume # Issue #

Received Date Month Year

Accepted Date Month Year

Published Date Month Year

Online at http://www.engj.org/

DOI:10.4186/ej.20xx.xx.x.xx

DOI:10.4186/ej.year.vol.issue.pp

1 Introduction

The measurement of blood volume is normally conducted *in vivo* using a technique called chamber-plethysmography. In order to conduct the experiment, the technique uses a chamber which makes the measurement need a plenty of space and complex measurement. However, in some cases, a simple technique that provides an immediate and advancement analysis is required. A non-invasive technique is used to obtain the blood volume changes called photoplethysmography (PPG). A PPG measurement device consists of an infra-red light source and a photo-detector. The infra-red light is illuminated through the skin. The penetration depth of light depends on the absorption and scattering coefficients of tissue. Therefore, the use of an infra-red light is preferred since it is more stable over time compared to red light [1]. The PPG instrument can be used in two modes, reflection mode and transmission mode. In the reflection mode, the light source and the photo-detector are placed side by side. The infra-red light is illuminated to the skin and the reflection light from tissue and bones is accepted by a photo-detector. Meanwhile in the transmission mode, the light is illuminated through the skin and detected by a photo-detector, which means that the light source and the photo-detector are facing each other. The light intensity accepted by the photo-detector represents the blood volume and the heart rate [2].

The PPG signal has a fundamental frequency of around 1 Hz. However, it has unequal periods. The PPG signal consists of pulsatile waveforms and DC components. The pulsatile waveforms related to the heart rate and the DC component represent the tissues and the average of blood volume [2]. A considerable number of studies on the processing of PPG signals using statistical and theory of non-linear dynamical analyses have been carried out [3, 4, 5, 6]. In addition, several time series analyses have increasingly been introduced to analyse the PPG signal, including fractal analysis. The fractal analysis provides an ability to measure the pattern behaviour of the time series and serial correlation over time windows [7]. Some of those fractal analyses introduce a Hurst exponent H which indicates the level of autocorrelated properties of the time series. One of these techniques is Detrended Fluctuation Analysis (DFA) [8].

There are increasing concerns about the differences of internal organ characteristics of men and women. Multiple factors have been cited by researchers to obtain the gender related differences of internal organs. [9] found that Body Mass Index (BMI) of men is higher than that of women. Using different approaches, numerous studies have attempted to explain the differences of the cardiovascular system of men and women, including Heart Rate Variability (HRV) or Heart Rate (HR). These previous studies have reported the relationships between the HRV or HR of ECG signals with the gender-related differences. All research findings indicate that there are significant differences of HRV parameters between men and women. [10] define that the high frequency heart rate fluctuation of men is greater than that of women in young subjects, and the heart rate dynamics of women is more complex than that of men. [11] also found that men in younger group (29-46 years old) had higher HRV parameters compared to women. Meanwhile there were no significant differences in the older group (64-74 years old). [12, 13] observed considerable differences between HRV parameters of men and women where the HRV parameters are higher in men. However, the features of PPG signal of men and women are still under explored. They have not been much investigated by researchers.

The HRV analysis is the most common analysis of cardiovascular system. PPG-HRV analysis was introduced as an alternative to analyse the PPG signal. [14, 15] applied HRV analysis to the PPG signal after pre-processing the PPG signal. Two features of HRV parameters (SDNN and rMSSD) could be extracted using this technique. However, according to the ECG-HRV analysis, only SDNN parameter value has a significant difference between men and women while there is no difference of rMSSD between men and women [12, 13]. Further research has reported that PPG-HRV technique generates unsatisfying results [16]. Previous studies have revealed the presence of fractal components in PPG signal [17, 18]. In particular, [19] applied DFA to PPG signal to study the hemodialysis of patients with diabetes mellitus by analysing the change of scaling exponent α. Fractal signal analysis has also been applied to a number of bio-signals to obtain important information, such as [20] who investigated the number of fractal component in human EEG signals using Coarse Graining Spectral Analysis (GCSA). [21] analysed EEG signals of 19 outpatients during hypnotising period using DFA. The use of DFA to discriminate EEG, EOG, EMG and ECG signals of sleep stages and sleep apnea severity was demonstrated by [22].

DOI:10.4186/ej.year.vol.issue.pp

Therefore, this study focuses on extracting more features of the PPG signal of the gender-related differences using PSD and DFA.

2 Methods

2.1 Instrumentation

Figure 1: The block diagram of PPG system which consists of infra-red sensor, filter and amplifier.

The PPG measurement was conducted using an infra-red sensor (HRM-215E) connected series with two stages of filter and amplifier. The sensor was connected to a High Pass Filter (HPF) with frequency cut off (*fc*) of 0.5 Hz to attenuate the DC component from the sensor and to avoid the PPG signal from going to the saturation mode. The output of HPF was connected to the first stage of active filter which consists of an op-amp based active Low Pass Filter (LPF) with frequency cut off of 3.4 Hz and gain (*A*) of 48. In order to pull down the swing output from the first stage, a similar HPF was placed before the second stage. The second stage filter and amplifier were connected to this HPF. Thus, the overall gain was approximately 2300. The last stage was an op-amp based buffer with a unity gain to match the impedance to the following instrument. Analog output of the PPG measurement device was connected to Arduino Uno for data acquisition with a transfer rate of 9600 bit per second. Arduino uno was connected to a laptop for saving the data as a csv (comma separate values) file for analysis using R language.

2.2 Subjects

The data acquisition involved two experimental groups of healthy man and woman volunteers who were between 20 and 25 years old since they provide strong emotional strain and less disease [23, 24]. Each group consists of 15 volunteers. During data acquisition, the two groups were seated comfortably and their left hand was bent down on a table with an angle of $130^o \pm 5^o$ in a room with a temperature of 26-28^{o}C. The data were recorded for 5 minutes as suggested by [25] for data processing using frequency domain method to the tip of the subject's left finger. Before the PPG pulse waves were acquired, the systolic and diastolic arterial blood pressure of the volunteers were taken. Subjects who smoked, had their period and took medications were excluded from this data acquisition. In this study, screening blood tests were not conducted. The procedure of data acquisition had been approved by the local Ethics Committee.

2.3 Heart Rate and Peak to Peak Interval

Heart rate (HR) is the number of heartbeat in a minute when contractions occur. The HR of adults when taking a rest is 60 to 100 beats per minute (bpm). Some factors that may affect HR are body position, body size and medication used. Heart rate is calculated using this formula :

$$bpm = \frac{Sampling\ rate}{(peak(i+1) - peak(i))} * 60; \qquad (1)$$

Peak to Peak Interval (PPI) is the width between two peaks as shown in Figure 2. The PPI represents cardiac beat-to-beat interval of PPG signal.

DOI:10.4186/ej.year.vol.issue.pp

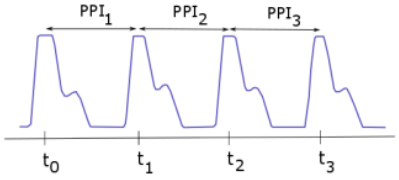


Figure 2: Peak to Peak Interval is the width between the peaks of PPG signal.

2.4 Power Spectral Density

In this study frequency domain analysis was conducted because in order to compute the parameters, it only needs short-time recording compared to time domain method [25]. To compute the frequency domain parameters, Power Spectral Density (PSD) analysis was used. PSD reveals the distribution of power signal over frequency, which can be defined as :

$$S_{xx}=\lim_{T\to\infty}\mathbf{E}[|x_T(\omega)|^2] \qquad (2)$$

where $\hat{x}_t(\omega)$ is Fourier transform of signal $x(t)$.

2.5 Detrended Fluctuation Analysis

DFA is a statistical technique for scaling long range correlations in a time series [8]. Consider, a 1-D time series X(i), i=1,...,N, where N is the length of the time series. The integrated time series can be computed using the following formula :

Figure 3: The method of DFA with a width of window. Similar method repeats for various width of window(w).

$$y(k)=\sum_{i=1}^{N}(X(i)-X_{avg}) \qquad (3)$$

where X_{avg} is the mean value of the time series X(i). The time series is divided into several windows (or segments) with width w. To determine the width of the windows w, the smallest value must be chosen to begin and the average fluctuation $F(w)$ of the signal is computed according to :

$$F(w)=\sqrt{\frac{1}{N}\sum_{k=1}^{N}(y(k)-y_n(k))^2} \qquad (4)$$

The process is repeated for all considered w and it ends on the largest desirable value like $N/2$ [7]. In each window, a double logarithmic graph, $log\ F(w)$ vs $log\ w$, is created to retrieve best approximation of data trend or linear approximation of the data between these two parameters by applying standard least square regression as shown in Figure 3. Hurst exponent (H) is a way to identify the autocorrelation in a time series. H values are divided by three categories. A value of 0.5 indicates an uncorrelated data or that the time series does not have previous value (no memory). Meanwhile, values of $0<H<0.5$ indicate that time series has a negative autocorrelation and values of $0.5<H<1$ indicate a positive autocorrelation of time series. The value of Hurst exponent describes the power-law correlation and the smoothness of time series [8].

DOI:10.4186/ej.year.vol.issue.pp

2.6 F-Test and T-Test

In order to determine the difference between the features of men and the feature of women, F-test and T-test (two tailed) were performed for each window. The tests were performed to analyse the difference of men and women features, from PSD and DFA computation. The following formula is used for T-test :

$$T = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{s_1^2/N_1 + s_2^2/N_2}} \quad (5)$$

where $\bar{Y}_1$ and $\bar{Y}_2$ are the sample means, s_1 and s_2 are the standard deviations of the samples and N_1 and N_2 are the sample sizes. The results of p-value from F-test indicates whether the variances of the two groups of data are equal or not. When the p-value > 0.05, it is assumed that the variances are equal and if the p-value < 0.05 then it is assumed that the variances are not equal. Based on the F-test, the T-test is conducted. When p-value of T-test < 0.05 then it is assumed that the data are different.

2.7 K-means Clustering

Clustering is a process to collect data that have similar characteristics into sets of groups. One of the clustering techniques is K-means clustering. K-means algorithm puts the data point in a dimensional space into K clusters. Each cluster is parameterized by its means ($\mathbf{m}^{(k)}$). K-means clustering algorithm is described as follows:

[Sorry. Ignored `\begin{algorithm} ... \end{algorithm}`]

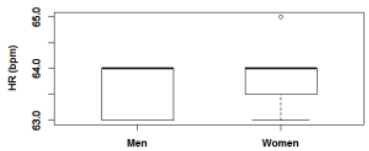
3 Results

Baseline characteristics in Table 1 reveal that there were notable significant differences between men and women in terms of systolic blood variable. The average of systolic blood pressure of men was 116.7 while the average blood pressure of women was 108. The difference is shown in the average of diastolic blood pressure, where the average of diastolic blood pressure for men was 83.3 and the average diastolic blood pressure for women was 77.3. This results align with previous research [10, 26, 27] which reported that women have lower blood pressure of about 5 to 10 mm Hg compared to men.

Table 1: Baseline characteristics of the subjects

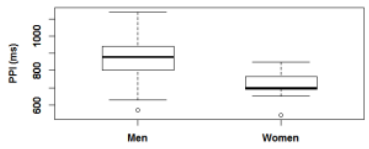
Variable	men	women
Ages	22 ± 1.5	23 ± 1.3
Systolic blood pressure (mm Hg)	116.7 ± 0.8	108 ± 0.5
Diastolic blood pressure (mm Hg)	83.3 ± 0.7	77.3 ± 0.5

3.1 Heart Rate and Peak to Peak Interval



DOI:10.4186/ej.year.vol.issue.pp

(a)



(b)

Figure 4: Box and whiskers plot of HR and PPI of volunteers

Figure 4 shows Box and Whiskers plot of HR and PPI of PPG signal of volunteers which are standard techniques for PPG features extraction. The plot shows that there was no significant difference between the HR of men and women. However, there was a significant difference between PPI features of men and women.

Table 2: Probability of HR and PPI features

Test	Heart rate	Peak to Peak Interval
p of F test	0.3562	0.0053
p of T test	0.3142	0.0032

F and T Test of HRV and PPI are shown in Table 2. The value of p of F test of HR was > 0.005 which means that the variances were equal. The p value of T test of HR was also >0.05 that means the HR features of men and women were not different. Different results were shown by PPI features in which p value of PPI features is < 0.005 which reveals that the PPI values between men and women were different.

3.2 Power Spectral Density

The technique divides spectral components into four ranges of frequency bands which are different from HRV parameters. Two samples of men and women were taken for ranges frequency band dividing analysis. The frequency band ranges have been chosen based on the analysis which was conducted by comparing two samples of PSD of men and women as shown in Figure 5. The spectral component ranges are Frequency Band 1 (0.5-0.8 Hz), Frequency Band 2 (0.8-1.3 Hz), Frequency Band 3 (1.3-1.8 Hz) and Frequency Band 4 (1.8-2.5 Hz). Frequency Band 1 was chosen since most of men spectral components are within this band. Meanwhile, most of women's spectral components were within Frequency Band 2. Frequency Band 3 was selected because within this band most of men's spectral components show higher magnitude compared to women. Finally, Frequency Band 4 was chosen to provide an additional band even though men's and women's spectral component within this band showing no significant differences.

(a)

(b)

Figure 5: Four ranges from two samples of PSD analysis of (a) men (b) women

DOI:10.4186/ej.year.vol.issue.pp

The four areas within the free frequency bands were computed . In order to obtain the features of the signals, the area relative (RA) to the total area was calculated as follows :

$$RA=(AFB/TA)*100; \quad (6)$$

where AFB is the area within the frequency band and TA is the total area within all frequency bands. The box and whiskers graph of the relative area is shown in Figure 6.

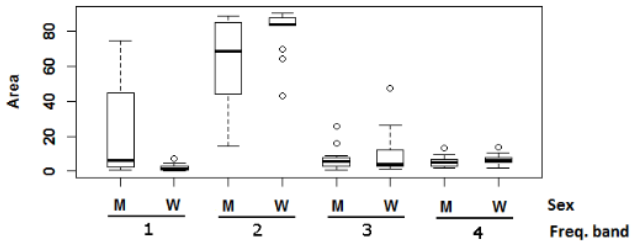


Figure 6: Average of relative area of 4 frequency bands of men (M) and women (W)

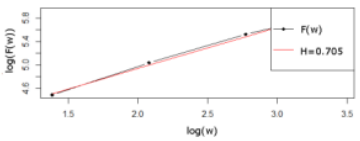
In order to obtain the difference of the result from PSD analysis, the F and T test (two tailed) were conducted and the result is shown in Table 3. It shows that variances of relative area within Frequency Band 1, Band 2 and Band 3 were unequal since p-values of F test < 0.05. Meanwhile, Frequency Band 4 had an equal variance. However, the T test showed that only Frequency Band 1 and Frequency Band 2 provided significant a difference where p-values < 0.05 and there was no significant difference for Frequency Band 3 and Band 4.

Table 3: Probability of features extraction from the PSD analysis

2***Test**	**Frequency bands**			
	1	**2**	**3**	**4**
p of F test	1.25E-13	0.005214	0.006430	0.4184930
p of T test	0.006070	0.023197	0.385169	0.2549640

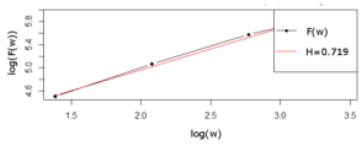
3.3 Detrended Fluctuation Analysis

The data acquisition was performed to each subject which produced 15 PPG time series signal for men and 15 PPG time series signal for women. Each time series of the signal was analysed using DFA. The DFA computation created 4 windows. The windows were multiple of two from 4 to 32 as shown in Figures 7(a) and 7(b).



(a)

DOI:10.4186/ej.year.vol.issue.pp



(b)

Figure 7: The example of one of the features of average fluctuation F(w) (a) men (b) women

Figure 8 shows the result of DFA computation of men (M) and women (W) for each window. The graphs show that for all windows, F(w) of women were higher than that of men. There were significant differences between the data of women and the data of men as the window increases. However, windows above 32 in this study were not considered because they generated almost similar features. Meanwhile, Hurst exponent result shows that the values were located between 0.5 and 1, which indicates that the DFA features of PPG signals of men and women had a memory or had a positive autocorrelation, as shown in Figure 9.

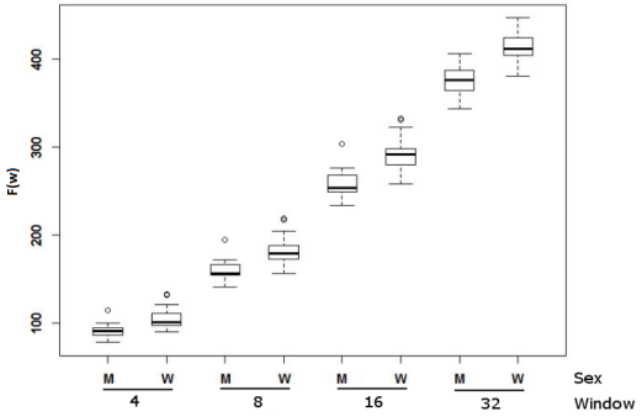


Figure 8: Average fluctuation F(w) of men (M) and women (W)

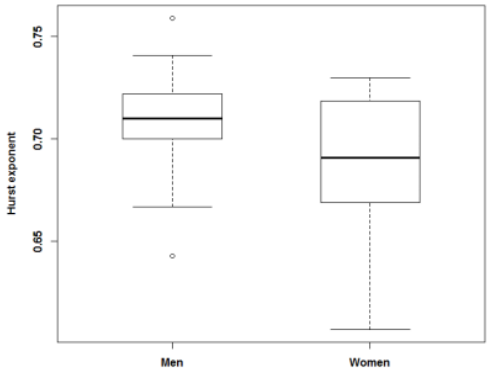


Figure 9: Box and whiskers graph of Hurst exponent of men's and women's features

DOI:10.4186/ej.year.vol.issue.pp

Table 4: Probability of features extraction from the DFA

2*Test	Width of windows (w)			
	4	8	16	32
p of F test	0.061545	0.092202	0.183118	0.335704
p of T test	0.001893	0.000830	0.000110	0.000003

Table 4 shows the p-value of F-test and T-test. The results of F-test show that the p-values on width of window 4,8,16 and 32 were bigger than 0.05 which reveals that the variances of men and women were equal on these windows. Based on the F-test, the T-test was performed, where p-values on all windows were smaller than 0.05. Bigger windows were not considered because the value of p of T-test > 0.05. Both tests reveal that the data of men and women were significantly different on short windows and provides best results for the feature extraction. The technique indicates that the pattern behaviour and serial correlation of features data series of men and women were different in short windows.

4 Discussion

Clinically there is no significant difference in terms of blood pressure of men and women. However, after puberty, men tend to have higher blood pressure. Furthermore, after menopause women tend to have higher blood pressure than men at that age [26]. Blood pressure is influenced by various factors such as heart as the main organ of blood circulation, age and gender. [10] found considerable differences between HRV parameters of men and women.

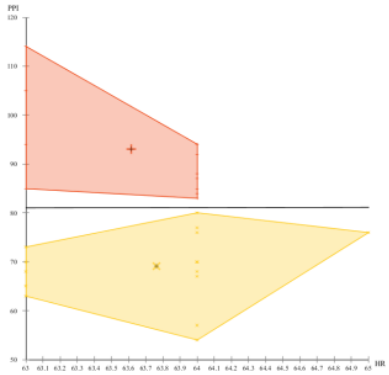


Figure 10: K-means clustering plot of HR and PPI features

K-means cluster analysis was used to evaluate the features of HR with PPI, PSD and DFA features using ELKI data mining software [28]. The result graphs are shown in Figure 10, 11 and 12. It shows that the features of HR with PPI, PSD and DFA were well separated into two groups of cluster memberships. The clustering result of HR with PPI is shown in Figure 10. It shows two memberships of groups for men and women and they were well separated. However, the plot also shows that HR features were identical between men and women features. The groups were separated by PPI features. Even though, the centroids were quite far from each other. Figure 11 shows the cluster memberships of PSD features. The plot shows the group membership for each band to others. It reveals that the features were grouped into two memberships. The centroids between two groups for each others were relatively far. However, there was an overlap of membership for Frequency Band 3 and Band 4. The final cluster centroids between the two groups were far.

DOI:10.4186/ej.year.vol.issue.pp

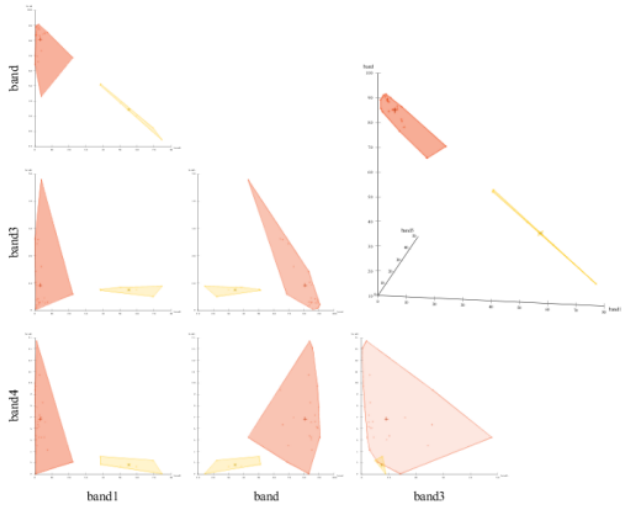


Figure 11: K-means clustering plot of PSD Features

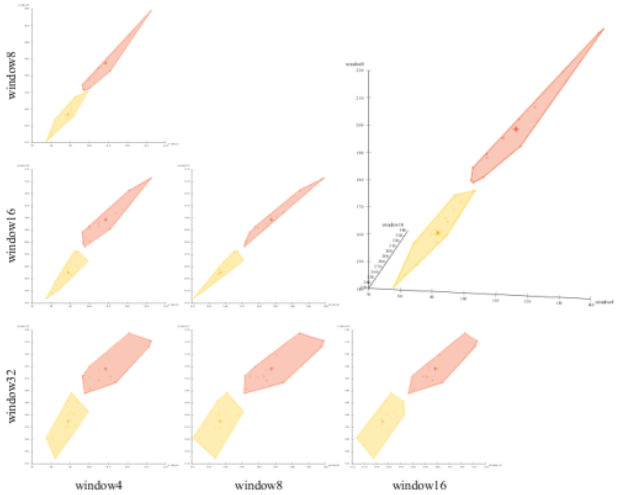


Figure 12: K-means clustering plot of DFA Features

The plot of K-means clustering for DFA features is shown in Figure 12. The plot shows results identical to the PSD features plot in which the DFA features were divided into the two groups. It demonstrates that the features for each window were well grouped into two memberships. The centroids between two groups for all occasions were significantly far. Overlapping membership was not shown for all groups. The expected result was also shown in the final clustering in which the features can be distinguished into the two groups memberships.

The results demonstrate that features from all techniques could be clustered. However, PSD and DFA techniques provide more features that can provide more accurate result.

5 Conclusions

This research was conducted to determine the difference of PPG signals of men and women. The signals were analysed using PSD and DFA. As the comparison used standard features techniques, HR and PPI

were calculated. The F-test and T-test were performed to distinguish the features of men and women for all techniques. There was no difference in HR features between men and women. Opposite result was shown by PPI features. The results reveal that there were significant differences features of men and women on Frequency Band 1 and Freqeuncy Band 2 using PSD technique and short windows (4,8,16 and 32) using DFA. To determine the difference of the data, F-test and T-test were conducted. The F and T test demonstrated that these two groups of data, men's and women's features were significantly different. A clustering analysis was applied to the features from all techniques to figure out whether the features can be distinguished into groups. K-means clustering was employed to collect similar features into groups. However, PSD and DFA techniques provided more features compared to HR with PPI. The results suggest that PSD and DFA are valuable methods for extracting the features of PPG signal and that the features could be used for classification purposes.

References

url@samestyle

[1] K. Shelley, "Photoplethysmography : Beyond the calculation of arterial oxygen saturation and heart rate," *Anesth. Analog*, vol. 105, pp. 1–6, 2007.

[2] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiological Measurement*, vol. 28, pp. 1–39, 2007.

[3] S. Flemming and L. Tarassenko, "A comparison of signal processing techniques for the extraction of breathing rate from the photoplethysmogram," *International Journal of Biological and Medical Sciences*, vol. 2, pp. 232–236, 2007.

[4] T. Fu, S. Liu, and K. Tang, "Heart rate extraction from photoplethysmogram waveform using wavelet multi-resolution analysis," *Journal of Medical and Biological Engineering*, vol. 29, pp. 229–232, 2008.

[5] J. Laźaro, E. Gill, R. Bailoń, and P. Laguna, "Deriving respiration from the pulse photoplethymosgraphic signal," *Computing in Cardiolog*, vol. 38, pp. 713–716, 2011.

[6] J. Lee and J. Nam, "Design of filter to reject motion artifacts of ppg signal by using two photosensors," *J. Inf. Commun. Converg. Eng.*, vol. 10, pp. 91–95, 2012.

[7] M. Riley, S. Bonnette, N. Kuznetsov, S. Wallot, and J. Gao, "A tutorial introduction to adaptive fractal analysis," *Frontiers in Physiology*, vol. 3, pp. 1–10, 2012.

[8] C. Peng, S. Havlin, H. Stanley, and A. Goldberger, "Quantification of scaling exponent and crossover phenomena in nonstationary heart beat time series," *Chaos*, vol. 5, pp. 82–87, 1995.

[9] A. Otte, J. Hassler, and J. Brogowski, "Relationship between body mass index and predicted %fat in college men and women," *Journal Health, Physical Education, Recreation & Dance*, vol. 10, pp. 23–29, 2000.

[10] S. Ryan, A. Goldberger, S. Pincus, J. Mietus, and L. Lipsitz, "Gender- and age-related differences in heart rate dynamics: Are women more complex than men?" *Journal of the American College of Cardiology*, vol. 24, pp. 1700–1707, 1994.

[11] P. Stein, R. Kleiger, and J. Rottman, "Differing effects of age on heart rate variability in men and women," *The American Journal of Cardiology*, vol. 80, pp. 302–305, 1997.

[12] D. Ramaekers, H. Ector, A. Aubert, A. Rubens, and F. V. de Werf, "Heart rate variability and heart rate in healthy volunteers. is the female autonomic nervous system cardioprotective ?" *European Heart Journal*, vol. 19, pp. 1334–1341, 1998.

[13] S. Saleem, M. Hussain, M. Khan, and S. Majeed, "Gender differences of heart rate variability in healthy volunteers," *Journal of Pakistan Medical Association*, vol. 62, pp. 422–405, 2012.

[14] M. Elgendi, M. Jonkman, and F. Boer, "Applying the apg to measure heart rate variability," in *Computer and Automation Engineering (ICCAE), 2010 The 2nd International Conference*, 2010.

[15] E. S. Hasan and J. Cáceres, "Heart rate variablity analysis based on photoplethysmographic signal," *Revista Cubana de Informatica Medica*, vol. 7, pp. 113–121, 2015.

[16] V. Jeyhani, S. Mahdiani, M. Peltokangas, and A. Vehkaoja, "Comparison of hrv parameters derived from photoplethysmography and electrocardiography signals," in *Proceedings of 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015, pp. 5952–5955.

[17] R. Enŕiquez, M. Castellanos, J. Rodríquez, and J. Caćeres, "Analysis of the photoplethymosgraphic signal by means of the decomposition in pricipal components," *Physiological Measurement*, vol. 23, pp. 17–29, 2002.

[18] =2plus 43minus 4 T. Yazawa, K. Tanaka, and T. Katsuyama, "Detrended fluctuation analysis of heartbeat interval signal: Alternans lowers the scaling exponent of heartbeat fluctuation dynamics in animal models and humans," in *Proceedings of the 2008 International Conference on BioMedical Engineering and*

Informatics - Volume 02, ser. BMEI '08. plus 0.5em minus 0.4emWashington, DC, USA: IEEE Computer Society, 2008, pp. 409–414. [Online]. Available: http://dx.doi.org/10.1109/BMEI.2008.314 =0pt

[19] Y. Tsuji, T. Asakawa, Y. Hitomi, A. Todo, T. Yoshida, and Y. Matsumoto, "Detrended fluctuation analysis of photoplethysmogarphy in diabetic nephropathy patiens on hemodialysis," *Brain and Health Informatics*, vol. 8211, pp. 218–224, 2013.

[20] M. Hazrati, A. Erfanian, and U. Hofman, "Fractal components from electroencephalogram provide features for brain computer interface," *Analysis of Biomedical Signals and Images*, vol. 20, pp. 387–393, 2010.

[21] J. Lee, D. Spiegel, S. Kim, J. Lee, S. Kim, B. Yang, J. Choi, Y. Kho, and J. Nam, "Fractal analysis of eeg in hypnosis and its relationship with hypnotizability," *Intl. Journal of Clinical and Experimental Hypnosis*, vol. 55, pp. 14–31, 2007.

[22] T. Penzel, J. Kantelhardt, H. Becker, J. Peter, and A. Bunde, "Detrended fluctuation analysis and spectral analysis of heart rate variability for sleep stage and sleep apnea identification," *Computers in Cardiology*, vol. 30, pp. 307–310, 2003.

[23] J. Kostis, A. Moreyra, M. Amendo, L. D. Pietro, N. Cosgrove, and P. Kuo, "The effect of age on heart rate in subjects free of heart disease," *Circulation*, vol. 65, pp. 141–145, 1982.

[24] E. Migliaro, P. Contreras, S. Bech, A. Extangibel, M.Castro, R. Ricca, and K. Vicente, "Relative influence of age, resting hear rate and sedentary life style in short-term analysis of heart rate variability," *Brazilian Journal of Medical and Biological Research*, vol. 34, pp. 493–500, 2001.

[25] Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, "Heart rate variability: Standards of measurement, physiological interpretation and clinical use," pp. 354–381, 1996.

[26] J. Reckelhoff, "Gender differences in the regulation of blood pressure," *Hypertension*, vol. 37, pp. 1199–1208, 2001.

[27] E. C. Pearce, *Anatomy and Physiology for Nurses*. plus 0.5em minus 0.4emFaber, 1966.

[28] E. Achtert, H.-P. Kriegel, and A. Zimek, " ELKI: A Software System for Evaluation of Subspace Clustering Algorithms," in *Scientific and Statistical Database Management*, ser. Lecture Notes in Computer Science, B. LudÃ¤scher and N. Mamoulis, Eds. plus 0.5em minus 0.4emSpringer Berlin Heidelberg, Jul. 2008, dOI: 10.1007/978-3-540-69497-7_41.

paper 5

ORIGINALITY REPORT

11%	5%	6%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

MATCH ALL SOURCES (ONLY SELECTED SOURCE PRINTED)

1%

★ "World Congress on Medical Physics and Biomedical Engineering 2018", Springer Nature, 2019

Publication

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	On		